

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:55:19 ; Search time 12.146 Seconds
(without alignments)
2841.441 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_359

Perfect score: 1773

Sequence: 1 GILVTVAPVAPVAPVAPVSGP.....KAGVTLHSVGPAAAGCTT 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	216	12.2	921	2 A48184	transcription init
2	216	12.2	921	2 A48184	transcription init
3	181.5	10.2	1275	2 T33369	hypothetical prote
4	176	9.9	2035	2 A40718	host cell factor C
5	173.5	9.8	528	2 A47141	gastric mucin (clo
6	169	9.5	3020	2 A43932	mucin 2 precursor,
7	167.5	9.4	549	2 C87719	protein R19.6 (im
8	165.5	9.4	825	2 T29836	hypothetical prote
9	165.5	9.3	2187	2 T30836	nascent polypeptid
10	164.5	9.3	752	4 A57784	AMUL/MTG8 mutant f
11	162	9.1	1367	1 S48478	glucan 1,4-alpha-g
12	158.5	8.9	1777	2 T34369	hypothetical prote
13	158	8.9	798	2 T34248	hypothetical prote
14	157	8.9	796	2 T21460	hypothetical prote
15	156	8.8	1324	2 S52863	DNA-binding protei
16	155.5	8.8	670	2 F36791	hypothetical prote
17	155.5	8.8	1151	2 T18535	high molecular mas
18	154.5	8.7	886	2 S29605	glycoprotein 350/2
19	154.5	8.7	3507	2 T34513	hypothetical prote
20	153.5	8.7	873	2 T44283	calphostin - fruit
21	153.5	8.7	2232	2 T34434	hypothetical prote
22	152	8.6	604	2 C57784	W68 protein splic
23	151	8.5	1169	2 S38181	floculation prote
24	150.5	8.5	1630	2 A53577	ascites stialoglyco
25	150	8.5	865	2 A47282	calcium-binding pr
26	150	8.5	873	2 F96615	probable Myb-famil
27	146.5	8.3	5376	2 T42215	zonadhesin - mouse
28	146	8.2	495	2 A44489	GT box-binding pro
29	145.5	8.2	3570	2 T45025	mucin MUC5B, trach

30	143.5	8.1	294	2 A37232	mucin, tracheal (A
31	143.5	8.1	797	1 VGBEX1	glycoprotein X pre
32	143.5	8.1	1547	2 T28657	blackjack protein,
33	143	8.1	1161	2 S57180	probable membrane
34	142.5	8.0	780	2 A48143	HF-1 regulatory el
35	142.5	8.0	2476	2 T34022	zonadhesin - pig
36	142	8.0	725	2 A41258	a agglutinin core
37	141	8.0	660	2 JMW067	chitinase (EC 3.2.
38	141	8.0	867	2 T45463	membrane glycoprot
39	140	7.9	1216	2 F88473	protein F40H6.5 (1
40	139	7.8	851	2 T22696	hypothetical prote
41	138.5	7.8	770	2 T22808	hypothetical prote
42	138.5	7.8	1032	2 T34433	hypothetical prote
43	138.5	7.8	1832	2 T31113	mucin-like glycopr
44	138.5	7.8	2422	2 T12687	ALR protein homolo
45	137.5	7.8	1002	2 A56678	yemanuclein-alpha

ALIGNMENTS

RESULT 1

A48184 transcription initiation factor IID 110K chain - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 16-Feb-1994 #sequence.revision 18-Nov-1994 #text.change 15-Oct-1999

C:Accession: A48184

R:Kokubo, T.; Gong, D.W.; Roeder, R.G.; Horikoshi, M.; Nakatani, Y.

A:Title: The Drosophila 110-KDa transcription factor TFIID subunit directly interacts

A:Reference number: A48184; MUID:93317591; PMID:8327460

A:Accession: A48184

A:Status: preliminary

A:Molecule type: mRNA; protein

A:Residues: 1-921 <KOK>

A:Cross-references: GB:563550; NID:9398432; PID:9398433

A:Experimental source: embryo nuclear extract

A:Note: Sequence extracted from NCBI backbone (NCBI:134863, NCBI:134864)

C:Genetics:

A:Gene: FlyBase:Tafl10

A:Cross-references: FlyBase:FBgn0010280

C:Keywords: transcription initiation

Query Match 12.2% Score 216; DB 2; Length 921;
Best Local Similarity 24.9% Pred. No. 4.7e-06;
Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;

QY	25	POIYAVKAPNTTIOFPANLQLPETVLIKNSGFLM-----VSPQOTVRAETTS	76
DB	133	POSPEITLSTLNTGTPA-----LLVKTNDGFLRVGTTGPPVTOTITWTSNNS	184
QY	77	NITSBPANPAPQVYKICTVNRSSQ-----LIKKVAVTPYKKAIOGTTVTVTP	127
DB	185	NTSTSTNHPPTTQ-TRLOTPAPAAASMTNTATSNITVSVSGVANSOPPHLTOLNAQ	243
QY	128	KPSVQSAVAPVTVTPGKPLNTVTTLKPSLGLASSTSPNEPLKAKENSAVAOINSP	187
DB	244	APQLPQITQITPAQOSOOQOVNVSSAGTATVSTTA-----ATT	287
QY	188	TMLENVK-KCKNFLMLIKLACSGSQSPKMGONKVKLEOLDAKIEAEFTRLKLYELK	246
DB	288	TOOGNTKEKCKRFLANLIEL--STREPKPVKEKNVTLLIOELVNAVNEDEECRDLERLIN	345
QY	247	SSPOHVLVPLIKSVVALROL-----LPSQSITQ---276	
DB	346	ASPOCLLGLFKSLPLRLQALYKELVIEGKPPQVHGLAGISQQLPIQAOIRIG	405
QY	277	---OCVOOTSSDMVIAICTTIVTTSPPVTTVSSSQSEKSIIVSGATAPRTVS---VQT	329
DB	406	PSQTTTIGQTVRM--TPNALGTPRPIGHTTTISKQPN---IRLPAPALVNTGIRI	460
QY	330	LNPLAGPVGAKAGVTLHSVGPAA 354	

DB 461 QIP-SLQVPGGANIVQIR--GPQHA 482

RESULT 2

A:Species: Drosophila sp.
C:Accession: A45183
C:Date: 10-Jun-1993 #sequence_revision 16-Nov-1994 #text_change 16-Feb-1997
R:Hoey, T.; Weinzler, R.O.; Gail, G.; Chen, J.L.; Dynlacht, B.D.; Tjian, R.
A:Title: Molecular cloning and functional analysis of Drosophila Taf110 reveal properties of the basal transcription machinery.
A:Reference number: A45183; MUID:93145326; PMID:7678780
A:Accession: A45183
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-921 <HOE>
A:Experimental source: embryo
C:Genetics:
A:Note: sequence extracted from NCBI backbone (NCBIP:123832)
A:Gene: FlyBase:Taf110
A:Cross-references: FlyBase:FBgn0010280

Query Match 12.2%; Score 216; DB 2; Length 921;
Best Local Similarity 24.9%; Pred. No. 4.7e-06;
Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;

OY 25 POLYAKPNTTTTIFPANLQLPCTVLKNSGFLM-----VSPOQTVAETTS 76
||:::||::| | | | | | | | | | | | | | | | | |
DB 133 PQSPITLTSTINTGTTPA-----LLVKTDNGFOLLRVGHTTGPPVTQTITWTSNS 184

OY 77 NITSRAVPANQVIKICTVPRNSSQ-----LIKKAVTPPKIAQIGTVYTVP 127
||:::||::| | | | | | | | | | | | | | | | | |
DB 165 NTTSTTNHPTTQ--TRIQTPPALASMTNTATSNIIIVSASSGVANSQDPHLTQLNAQ 243

OY 128 KPSSQSAVAPRTSVTVTPGKRPLNVTTLKPSSLGASTPSSENPYLKAENSAVAQINSP 187
| | | | | | | | | | | | | | | | | | | | | | | |
DB 244 APOLQIQTQIQTPIPAQSOQQOVNVNVSAGTATAVSSTA-----ATT 287

OY 188 TMLENVK-KCKNFLMLIKLACSGSQSPDMQNKKLEOLDLAKIEAEFEFTRLKYVELK 246
| | | | | | | | | | | | | | | | | | | | | | | |
DB 288 TOQGNTKEKCRKFNLNIEL--STRPKPVKENVNTLIQELVNANVEEFECRDRLERLN 345

OY 247 SSPQHLVPELKKSVALROL-----LPMSOSFIQ--- 276
| | | | | | | | | | | | | | | | | | | | | | | |
DB 346 ASPQCLGLGFLLKSLPLRLALYTKELYEGIKRPPOHYLAGLSQQLPKIOAOIRPFG 405

OY 277 ---QCVOOTSADVIATCTTYYTSPVYTTYSSSSOSEKSIISGATAPRTVS---VQT 329
| | | | | | | | | | | | | | | | | | | | | | | |
DB 406 PSQTTTIQOTQVRMI--TPNALGTPRPPIGHITTISKQPN--IKLPAPRLVANGIRT 460

OY 330 LNPFLAGPVGAAGVYTLHSVGTAA 354
| | | | | | | | | | | | | | | | | | | | | | | |
DB 461 QIP-SLQVPGGANIVQIR--GPQHA 482

RESULT 3

T33369
hypothetical protein H02F09_3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2000
C:Accession: T33369
R:Geisel, C.; Harmon, G.
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1275 <GEI>
A:Cross-references: EMBL:AF077538; PIDN:AAC64622.1; GSPDB:GN00028; CESP:H02F09.3
A:Experimental source: strain Bristol NZ; clone H02F09
C:Genetics:

A:Gene: CESP:H02F09.3
A:Map position: X
A:Introns: 42/1; 78/3; 106/3; 135/2; 161/2; 224/1
C:Superfamily: yeast glucan 1, α -glucan-glucosidase homolog; glucan 1, α -glucosida
Query Match 10.2%; Score 181.5; DB 2; Length 1275;
Best Local Similarity 23.1%; Pred. No. 0.0011;
Matches 93; Conservative 62; Mismatches 151; Indels 97; Gaps 16;

OY 2 TLTVKAVASAPKVSQSP---RLPAQIVAKAPNTTTIQFPANLQLPPTGLVLIKS- 56
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 275 TTVTPTTITGPTVVVTVPITVTVTPSTVPMSPITPTSTVTVTPSVTAFTKRPST 334
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY 57 ---SGCLMAY-V-PQQVTAETTSNITSRPAVPANPDYKICTVNRSSQLIKKAATPVK 113
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 335 VVTASTVTVTPSTVTVTPNTV--VTSSPTAAATPTTV--VTTPST-----VTVTPS 382
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY 114 KLAQIGTIIVV-----TTVKPSVSSQSAVPTGSVNT-----VTPEKPLNTV 153
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 363 TVVIVPTTVTNPSVTVAPSTV--VIYPVTVMHSRSRTVITPTTGSGSSPTACTSLAST 440
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY 154 TTLKPSLSGASTSPSEPNLKAENSAVAOVINLSPTMLENVKCKNFLAMLIKLAGSSQS 213
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 441 AVTTETSIGSSSTP-----LPQSGTSLMSISLTSTYTPSS-----STAAGATS 481
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY 214 PEMGNVKKIIEQLDA--KIIEEFTRKLYELKSSPHQLVLPFLKKSVALLQLPNS 271
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 482 PATQGSTPTGTGSMSSGPTTVAPGASTESTVLQSTPGCTV-----TLPSG 529
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY 272 OSFIQCQVOQTSDPVIANTCCTTITVTSPVNTTVSSSQSEKSIIVSGATAADPRYSVOTLN 331
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 530 SSTAFAGISPOASTYTWTDTISTVSGSTVTSQTAESSLSTSPISAGSST-STVSTVSSQ 588
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY 332 P-----LAGPVGAKAGVTLHSGPTAATGGTT 359
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 589 PSTVIPVASASIYSTLGSGTGA-----SPGTESGSGST 625
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 4
A40718
host cell factor C1 precursor - human
N:Alternate names: CFF; HCF; VP16 accessory protein host cell factor (VCAR)
C:Species: Homo sapiens (man)
C:Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text-change 20-Apr-2000
A:Accession: A40718; GenBank: U02511; EMBL: AF0086; JGI: 137453
R:Wilson, A.C.; Lamarco, K.; Peterson, M.G.; Herr, W.
Cell 74, 115-125, 1993
A>Title: The VP16 accessory protein HCF is a family of polypeptides processed from a
A:Reference number: A40718; MUID:93327419; PMID:8392914
A:Accession: A40718
A>Status: Preliminary; not compared with conceptual translation
A:Molecule type: mRNA; protein
A:Residues: 1-2035 <WIL>
A:Cross-references: PIDN:AAB27583.1; PID:g399752
A:Experimental source: HeLa cell
A>Note: sequence extracted from NCBI backbone (NCBIP:135349)
R:Platzter, M.; Bauer, D.; Drescher, B.
submitted to the EMBL Data Library, March 1995
A:Reference number: H013368
A:Accession: G02511
A>Status: translated from GB/EMBL/DDBT
A:Molecule type: DNA
A:Residues: 1-2035 <PLA>
A:Cross-references: MID:g1302657; PIDN:AAC51751.1; PID:g1302663
R:Kristle, T.M.; Pomerantz, J.L.; Twomey, T.C.; Parent, S.A.; Sharp, P.A.
J Biol. Chem. 270, 4387-4394, 1995
A>Title: The cellular C1 factor of the herpes simplex virus enhancer complex is a fam
A:Reference number: A56088; MUID:95181425; PMID:7876203
A:Accession: A56088
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-664; 'T', 666-1637, 'E', 1639-1684, 'A', 1686-1734, 'Q', 1736-2035 <KR>
R:Reitlin, A.; Miranda, S.; Redolfi, E.; Zucchi, I.; Villa, A.; Patrosso, M.C.; Stri

A:Reference number: A91831; MUID:87194600; PMID:3106330
 A:Accession: A26877
 A:Molecule type: DNA
 A:Residues: 1-242 <YAM>
 A:Cross-References: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525
 A:Accession: B26877
 A:Molecule type: DNA
 A:Residues: 762-1331 <YAZ>
 A:Cross-References: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526
 R:Parlo, J.M.; Ianez, E.; Zalcacain, M.; Claros, M.G.; Jimenez, A.
 FEBS Lett. 239, 179-184, 1988
 A>Title: Similar acid elements in the 5' regions of the STRA and SGA genes from Sacchar
 A:Reference number: S27281; MUID:89031230; PMID:3141213
 A:Accession: S27281
 A:Molecule type: DNA
 A:Residues: 1-31 <PAR>
 A:Cross-References: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
 R:Lambrecht, M.G.; Bauer, F.F.; Marmur, J.; Precorius, I.S.
 Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
 A>Title: Muc, a mucin-like protein that is regulated by Mas10, is critical for pseudohy
 A:Reference number: JG6123; MUID:96323237; PMID:8710886
 A:Accession: JG6123
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-1367 <LAM>
 A:Cross-References: GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387
 C:Genetics:
 A:Gene: SGD:WUC1; STRA2; MAL5; DEX2; SGD:S0001458
 A:Cross-References: MIPS:YIR019C; SGD:S0001458
 A:Map position: 9R
 C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
 F:5-21/Domain: transmembrane #status predicted <TML>
 F:1350-1366/Domain: transmembrane #status predicted <TM2>
 Query Match 9.1%; Score 162; DB 1; Length 1367;
 Best Local Similarity 21.2%; Pred. No. 0.021;
 Matches 87; Conservative 62; Mismatches 181; Indels 80; Gaps 13;
 QY 2 TLVTKVAPVSAAPKVSQPRLPAP-----QIVAAKAPNTTITQIPANLQIPGTVLIR 54
 DB 618 TTESSAPVTSSTESSAPVPTPSSSTESSAPVPTPSSSTESSAPVPTPSSSTTE 677
 QY 55 SNGSPMLVSPQOT--VTRAETSNITSRPNANPQVTKICTVNSSQLKKYAV-- 109
 DB 678 SSSAPVTSSTESSAPVTSSTESSAPVPTPSSSTESSAPVPTPSSSTESSAPV 737
 QY 110 -TPVKIAOIGTVVT-----VPRSSVQS-----VAPTSVTVTPGKPLNT- 152
 DB 738 PTPSSSTESSAPVTSSTESSAPVPTPSSSTESSAPVPTPSSSTESSAPVPTP 797
 QY 153 -----VTTLKPSLGAASSTPNEPNLKAENSAVQINLSPTMLENVKCKNFIA 201
 DB 798 SSSSTESSAPVPTPSSSNITSSAPSTPSSSTESSAPV--PTPSSSTESS-- 851
 QY 202 MLTKLACSSQS-----PEMQONKILVEQLDAKIAEETRLIYELKSSPPHLY 254
 DB 852 --APVSSSTESSAPVPTPSSSNITSSAPSSIPSSSTESSSTGTTPSSSK----- 904
 QY 255 PFLKSVVALROLPLNSQFIOCVQOTSSDMVIACTTCTVTTPSP--VTTTVSSQSE 311
 DB 905 -----YRGSQF--ETSVSTETETIVPTKITTSTVPTTTTITTTTTCSTGIN 949
 QY 312 KSIIVSGATAPRTV--SVQTLNPLAGPVGAKAGVTLHSGVPTAATGGTT 359
 DB 950 SAGETSGCSPKVTVTVTPTTTTSTTTTSTTTTTCSTGNSAGETT 999
 RESULT 12
 T34369
 hypothetical protein T19D12.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34369
 R:Favell, A.
 submitted to the EMBL Data Library, November 1995
 A:Description: The sequence of C. elegans cosmid T19D12.
 A:Reference number: 221513
 A:Accession: T34369
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1777 <FAV>
 A:Cross-References: EMBL:U41263; PIDN:AAC24428.1; GSPDB:GN00020; CESP:T19D12.1
 A:Experimental source: strain Bristol N2; clone T19D12
 C:Genetics:
 A:Gene: CESP:T19D12.1
 A:Map position: 2
 A:Introns: 36/1; 134/2; 180/1; 622/3; 691/2; 754/1; 1111/2; 1174/1; 1271/3; 1322/2; 1
 Query Match 8.9%; Score 158.5; DB 2; Length 1777;
 Best Local Similarity 24.0%; Pred. No. 0.049;
 Matches 94; Conservative 51; Mismatches 158; Indels 89; Gaps 17;
 QY 3 LVTKVAPVSAAPKVSQPRLPAPQIVAAKAPNTTITQIPANLQIPGTVLIRKNS--GP 59
 DB 222 VTTTVAQTAPLIVTTAN--TTQGVTTTACGVTTVYRAQNSFLAFTTAPSTNTTQCV 279
 QY 60 LMLVSPQOITRAE-----TTSNITSRPANPQVTKICT-----VPNS-----S 101
 DB 280 TTTVGKTTVTTAQNSTMAATTTASNTTQPVVTTSTSTGISTTTAQAAPPSSSVIPTTT 339
 QY 102 QLIKKAIVAPVKKLAOIGTVVTVTPKPSVO--SVAVP--TSVAVTTPGKPLNTVTLIK 157
 DB 340 QTTQRPSTICIPSTVSTSCSTSTSPSTQYSSSPSTYSNFTTPTTLLT-STIA 398
 QY 158 PSSIGA--STSPNEPNLKAENSAVQINLSPTMLENVKCKNFLAMLIKACGSSQSPM 216
 DB 399 PSTQGVPTSSKSSPN-----STPTTTTTP-----GASSSTL 430
 QY 217 GQVKKLVEQLDAKIAEETRLKLYELKSSPPHLYPFL-----KSVVALROL 268
 DB 431 GSSSTSTVSTTTPS-----TPKVSSTLTSQSPPTSPPLVSSSSGSSSTVTVSTIT 484
 QY 269 PNSQSF--IOCVQOTSSDMVIACTTCTVTTPSPVTTTVSSSQSEKSLIYGATAPRTVS 327
 DB 485 PSTQGVPTSTSNQPTPSTSTPTPKSTVANS--STTGATSTASPTTITTSAPTSSHP 542
 QY 328 QTLNPLAGPVGAKAGVTLHSGVPTAATGGTT 359
 DB 543 SS-----TMTSTVPTSTPAST 559
 RESULT 13
 T34248
 hypothetical protein F31D5.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T34248
 R:Milcox, L.
 submitted to the EMBL Data Library, June 1995
 A:Description: The sequence of C. elegans cosmid F31D5.
 A:Reference number: 221494
 A:Accession: T34248
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-798 <MIL>
 A:Cross-References: EMBL:U28941; PIDN:AAC71100.1; GSPDB:GN00020; CESP:F31D5.4
 A:Experimental source: strain Bristol N2; clone F31D5
 C:Genetics:
 A:Gene: CESP:F31D5.4
 A:Map position: 2
 A:Introns: 63/1; 95/2; 122/3; 189/2; 309/2; 350/3; 364/3; 411/2; 414/1; 443/1; 543/3;
 Query Match 8.9%; Score 158; DB 2; Length 798;
 Best Local Similarity 23.8%; Pred. No. 0.019;
 Matches 79; Conservative 46; Mismatches 125; Indels 82; Gaps 14;

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:54:50 ; Search time 6.91863 Seconds

(without alignments)
2152.161 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_359

Perfect score: 1773
Sequence: 1 GTLVTKVAPVSAAPKVVSSGP.....KAGVTILHSVGPAAAGCTT 359

Scoring table: BLOSUM62
Gap 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1773	100.0	1 T2DT_HUMAN	092750 homo sapien
2	491	27.7	1 T2D3_HUMAN	000268 homo sapien
3	216	12.2	1 T2D3_DROME	P47823 DROSOPHILA
4	177.5	10.0	1 HFC1_MESAU	P51611 mesocricetu
5	175	9.9	1 HFC1_HUMAN	P51610 homo sapien
6	167.5	9.4	1 MUC2_HUMAN	002817 homo sapien
7	162	9.1	1 AMYH_YEAST	P08640 saccharomyc
8	157	8.9	1 V630_HSV1	009624 caenorhabdi
9	155.5	8.8	1 V630_HSV1	000130 ictaluriid h
10	154.5	8.7	1 V6P3_EBYA8	007284 epstein-bar
11	153	8.6	1 SP2_HUMAN	002086 homo sapien
12	152	8.5	1 MTG8_HUMAN	006455 homo sapien
13	151	8.5	1 YK82_YEAST	P36170 saccharomyc
14	150	8.5	1 CPN_DROME	002910 drosophila
15	147	8.3	1 ZAN_MOUSE	Q61909 mus musculu
16	146.5	8.3	1 ZAN_MOUSE	088799 mus musculu
17	145.5	8.2	1 PODX_HUMAN	000592 homo sapien
18	143.5	8.1	1 VGLX_HSVB	P28968 equine hebp
19	143	8.1	1 DAN4_YEAST	P47179 saccharomyc
20	142.5	8.0	1 ZAN_PIG	028983 sus scrofa
21	142	8.0	1 AGAL_YEAST	P32323 saccharomyc
22	140	7.9	1 YPX5_CAEEL	009277 caenorhabdi
23	139.5	7.9	1 AL51_CANAL	P46590 candida alb
24	137.5	7.8	1 YEM4_DROME	P25992 drosophila
25	137	7.7	1 YJH8_YEAST	P47033 saccharomyc
26	136.5	7.7	1 CH12_COCIM	P54197 coccidioidi
27	135.5	7.6	1 OCT1_PIG	Q29076 sus scrofa
28	135.5	7.6	1 P3K2_DICDI	P54674 dictyosteli
29	134.5	7.6	1 VGP3_EBY	P03200 epstein-bar
30	134	7.6	1 FLOS_YEAST	P38894 saccharomyc
31	133.5	7.5	1 MUC1_XENLA	005049 xenopus lae
32	132.5	7.4	1 YM96_YEAST	004893 saccharomyc
33	132	7.4	1 AL53_CANAL	074623 candida alb

34	131	7.4	1036	1 P200_MYCPN	P75211 mycoplasma
35	128.5	7.2	743	1 OCT1_HUMAN	P14859 homo sapien
36	128	7.2	3866	1 HRX_MOUSE	P55200 mus musculu
37	127.5	7.2	1574	1 SYJ1_RAT	Q62910 rattus norv
38	127	7.2	634	1 HWP1_CANAL	P46593 candida alb
39	127	7.2	1306	1 MSB2_YEAST	P32334 saccharomyc
40	127	7.2	2700	1 ZAN_HUMAN	Q39493 homo sapien
41	126.5	7.1	784	1 SP4_HUMAN	Q02446 homo sapien
42	125.5	7.1	1794	1 YAV1_SCHPO	Q10172 schizosach
43	125	7.1	1048	1 P100_HCMVA	P08318 human cytom
44	125	7.1	1322	1 YAG3_YEAST	P39712 saccharomyc
45	124.5	7.0	1147	1 SRE1_HUMAN	P36956 homo sapien

ALIGNMENTS

RESULT 1
T2DT_HUMAN STANDARD; PRF; 801 AA.

AC Q92750;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription Initiation factor TFIID 105 kDa subunit (TAFII-105)
DE (TAFII105) (Fragment).
GN TAF4B OR TAF2C2 OR TAFII105.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=97011146; PubMed=8858156;
RA Dikstein R., Zhou S., Tjian R.;
RT "Human TAFII 105 is a cell type-specific TFIID subunit related to hTAFII130.";
RL Cell 87:137-146(1996).

CC -1- FUNCTION: CELL TYPE-SPECIFIC SUBUNIT OF TFIID THAT MAY FUNCTION AS A GENE-SELECTIVE COACTIVATOR IN CERTAIN CELLS. TFIID IS A MULTIMERIC PROTEIN COMPLEX THAT PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS ACTIVATORS AND REPRESSORS.
CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A NUMBER OF TBP-ASSOCIATED FACTORS (TAFs). TAFII105 MAY FORM HETERODIMERS WITH TAFII130.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: FOUND IN B CELLS BUT NOT IN OTHER TESTED CELLS.

CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC EMBL: Y09321; CAA70499.1; -
CC Genew: HGNC:11538; TAF4B.
CC MIM: 601689; -
CC DR InterPro: IPR003894; TAF_hom.
CC DR SMART: SM00549; TAFH; 1.
CC FT Transcription regulation; Nuclear protein.
CC NON TER 1
CC SEQUENCE 801 AA: 85658 MW; D12B4932EFA9CD2 CRC64;

Query Match 100.0%; Score 1773; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 2.2e-100;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 1 GTLVTKVAPVSAAPKVVSSGPRLPAQIVAKAPNTTTOFPANLQLPFGYVLIKNSGPL 60
|||||

```

Db 1 GRLVTVAPVAPSAAPKSSSGRLPAPOLIVANKANTTTTIOFPANLOLPFGTVILKNSGRL 60
QY 61 MLVSPQOQVTRAFETNTSITSRPAVPANPQVTKICTVPSNSSQLIKKAVATPVKKLAQIGT 120
Db 61 MLVSPQOQVTRAFETNTSITSRPAVPANPQVTKICTVPSNSSQLIKKAVATPVKKLAQIGT 120
QY 121 TVVTVTPKSSVQSVAVPVSVTVTPGKPLNTVTTTLKPSLSGASPPSNPNKAKESNA 180
Db 121 TVVTVTPKSSVQSVAVPVSVTVTPGKPLNTVTTTLKPSLSGASPPSNPNKAKESNA 180
QY 181 VOINLSPTMLENVKCKNFAMLIKILACSSQSPENQONKLVLEQLDAKIAEETFRK 240
Db 181 VOINLSPTMLENVKCKNFAMLIKILACSSQSPENQONKLVLEQLDAKIAEETFRK 240
QY 241 LVELKSSPQPHLVPLFLKSSVALROLPLNSQSFIOCCVOQISSDMVIATCTTTVTTSVP 300
Db 241 LVELKSSPQPHLVPLFLKSSVALROLPLNSQSFIOCCVOQISSDMVIATCTTTVTTSVP 300
QY 301 VTTTVSSQSEKSIIVSGATPRTVSQVOTLNLAPGVGAKGVVTLHSVCPATAGTGT 359
Db 301 VTTTVSSQSEKSIIVSGATPRTVSQVOTLNLAPGVGAKGVVTLHSVCPATAGTGT 359

RESULT 2
T2D3_HUMAN STANDARD: PRT; 1083 AA.
ID T2D3_HUMAN STANDARD: PRT; 1083 AA.
AC 000268; Q98721; Q98X42; Q98R40;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription initiation factor TFIIID 135 kDa subunit (TAFII-135)
DE (TAFII135) (TAFII130) (TAFII130).
GN TAF4 OR TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-97336072; PubMed-9192867;
RA Mengus G., May M., Carre L., Chamdon P., Davidson I.;
RT "Human TAF(II)135 potentiates transcriptional activation by the AF-2s
RT of the retinoic acid, vitamin D3, and thyroid hormone receptors in
RT mammalian cells."
RL Genes Dev. 11:1381-1395(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-21638749; PubMed-11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Graham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kershley A.M., King A., Knighs C., Lloyd D.M., Lovell J.D.,
RA Levasalho M.H., Leverisha M.A., Lloyd C., Lloyd D.M., McMurtry A.J.,
RA Marsh V.L., Martin S.L., McConachie L.J., McElay K., McMurtry A.J.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plunk R.W., Ramsay H.,
RA Rice C.M., Smith M.L., Soderlund C., Stewart C.A., Sulston J.E.,
RA Skuce C.D., Smit M.L., Soderlund C., Stewart C.A., Sulston J.E.,
RA Swann R.M., Symons N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin R., Wall M., Wallis J.M.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilmington L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;

```

```

RT "The DNA sequence and comparative analysis of human chromosome 20."
RL Nature 414:865-871(2001).
RN [3]
RP SEQUENCE OF 105-1083 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE-97098442; PubMed-8942982;
RA Tanese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;
RT "Molecular cloning and analysis of two subunits of the human TFIIID
RT complex: hTAFII130 and hTAFII100."
RL Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).
CC -1- FUNCTION: MAKES PART OF TFIIID IS A MULTIMERIC PROTEIN COMPLEX THAT
CC PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS
CC ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION
CC BY THE AF-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.
CC SUBUNIT: TFIIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE TAF6 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Y11354; CAA72189.1; -
CC EMBL: AL137077; CAC36006.1; -
CC EMBL: AL109911; CAC22312.2; -
CC EMBL: U75308; AAC50901.1; -
CC TRANSFAC: T02328; -
CC Genew; HGNC:11537; TAF4.
CC MIM: 601796; -
CC InterPro: IPR003894; TAF_hom.
CC SMART: SM00549; TAFH; 1.
CC Transcription regulation; Nuclear protein.
CC
CC FT DOMAIN 39 42 POLY-HIS.
CC FT DOMAIN 52 57 POLY-ALA.
CC FT DOMAIN 98 101 POLY-GLY.
CC FT DOMAIN 142 148 POLY-ALA.
CC FT DOMAIN 268 275 POLY-PRO.
CC FT DOMAIN 331 337 POLY-ALA.
CC FT DOMAIN 680 683 POLY-PRO.
CC FT DOMAIN 808 813 POLY-ALA.
CC FT DOMAIN 828 831 POLY-ASP.
CC FT DOMAIN 828 831 POLY-ASP.
CC FT CONFLICT 105 117
CC FT CONFLICT 136 136 A -> S (IN REF. 2).
CC FT CONFLICT 185 185 G -> GPG (IN REF. 2).
CC FT CONFLICT 233 264 MISSING (IN REF. 3).
CC FT CONFLICT 293 293 P -> L (IN REF. 3).
CC SEQUENCE 1083 AA; 109943 MW; A6453827572A0752 CRC64;

Query Match 27.7%; Score 491; DB 1; Length 1083;
Best local Similarity 38.4%; Pred. No. 1.4e-22;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;
QY 13 PKVSSG-----PRLPAPQIVAVKAPVTTTIOFPANLOLPFGTVILKNSGRL 61
Db 414 PRLPAPQIVAVKAPVTTTIOFPANLOLPFGTVILKNSGRL 61
QY 62 LVSPQOQVTRAFETNTSITSRPAVPANPQVTKICTVPSNSSQLIKKAVATPVKKLAQIGT 115
Db 465 MI-POALOMQAOAHQAPQTPMAPPAPPSAPVOISTVQAPGPIIAR-QVTP----- 518
QY 116 AQGTIVTVTPKSSVQSVAVPVSVTVTPGKPLNTVTTTLKPSLSGASPPSNPNKAKESNA 168
Db 519 ----TIIKOV-----SAQGTVPQSAFLQSPGVQPOLVLGAAGTSLGATVAVQGTGTPQ 571
QY 169 NE-PNKAENSAVQINISPTMLENVKCKNFAMLIKILACSSQSPENQONKLVLEQLDAKIAEETFRK 227
Db 572 RTVPGATTTISSATE-----TMENVKCKNFAMLIKILACSSQSPENQONKLVLEQLDAKIAEETFRK 625

```


QY 228 LDATIEBETFRKLYELKSSPOPHLYPFLKKSVALROLIPNSQSFIOQCVOO---TS 283
 DB 626 LDGIEEDFTSRILRSLNSPOPYLVPFLKRSIPALROLPPDAASAIIOQOQPPPTS 685
 OY 284 SDWIACTTTTSPVTTTSSSQSEKSTIVSGATAPRTVSV 327
 DB 686 Q-----ATTALVAVLSSSVORTACKTAVTYSALQPVLSL 722
 RESULT 3
 T203 DROME STANDARD; PRT: 921 AA.
 AC P47825; P49845; Q9YU7;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcription initiation factor TFIID 110 kDa subunit (P110)
 GN TAF110 OR GC5444.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID:7227;
 RN 111
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 519-540; 597-616 AND 857-874.
 RC TISSUE-Embryo;
 RX MEDLINE-93145326; PubMed-7678780;
 RA Hoeij T., Weinzler R.O.J., Gill G., Chen J.-L., Dynlacht B.D.,
 RT "Molecular cloning and functional analysis of Drosophila TAF110
 RL reveal properties expected of coactivators.";
 RT Cell 72:247-260(1993).
 RT Proc. Natl. Acad. Sci. U.S.A. 90:5896-5900(1993).
 RN 131
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 398-406; 520-540 AND 860-877.
 RC TISSUE-Embryo;
 RX MEDLINE-93317591; PubMed-8327460;
 RA Kokubo T., Gong D.-W., Roeder R.G., Horikoshi M., Nakatani Y.,
 RT "The Drosophila 110-kDa transcription factor TFIID subunit directly
 RL interacts with the N-terminal region of the 230-kDa subunit.";
 RT Proc. Natl. Acad. Sci. U.S.A. 90:5896-5900(1993).
 RN 131
 RP SEQUENCE FROM N.A.
 RC STRAIN-Berkeley;
 RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,
 RA Botkova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
 RA Burks K.C., Busam D.A., Butler H., Cadiot E.E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davaport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson R., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harit N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Sytkes R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveli J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: TFIID IS A MULTIMERIC PROTEIN COMPLEX THAT PLAYS A
 CC CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS ACTIVATORS
 CC AND REPRESSORS. MAY FUNCTION AS A COACTIVATOR BY SERVING AS A SITE
 CC OF PROTEIN-PROTEIN CONTACT BETWEEN ACTIVATORS LIKE SPI (OR BTB)
 CC AND TFIID COMPLEX.
 CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
 CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L06861; NOT_ANNOTATED_CDS.
 CC EMBL; S63550; AAB27433.1;
 CC EMBL; AE003528; AAF49536.1;
 CC TRANSFAC; T02121;
 CC FLYBASE; FBgn0010280; Tafi110.
 CC InterPro: IPR0003894; TAF_hom.
 CC SMART; SM00549; TAFH; 1.
 CC
 CC TRANSCRIPTION regulation; Nuclear protein.
 CC DOMAIN 66 82 POLY-GLN.
 CC FT DOMAIN 108 111 POLY-GLN.
 CC FT DOMAIN 259 265 POLY-GLN.
 CC SO SEQUENCE 921 AA; 99338 MW; 27E6852659872767 CRC64;
 Query Match 12.28; Score 216; DB 1; Length 921;
 Best Local Similarity 24.9%; Pred. No. 5,4e-06;
 Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;
 OY 25 PQIAYVAKPNTTTOFPPNQLPPGTVLKNSGPIML-----VSPOQTVRAERTS 76
 DB 133 PQSPTSLTSTANTGTPA-----LVKTDNGQLRLVGTGGPPVTOTITNTSNN 184
 OY 77 NITSRPAVPANPQIVKICTVFNSSSQ-----LIKKVAVTPVKILQIGTVVTVTP 127
 DB 185 NITSTTNPTTQ-IRQTVPAAASMTNTATSNITVNASVSGANSQPHLTQLNAQ 243
 OY 128 KPSSVQSAVPTSVVYTPGKPLNTVTTLKPSISLGASSTPSENEPLKAENSAVAQINISP 187
 DB 244 APOLPQITQITQIPAOOSQOQVNNVSSAGCTAAVASTTA-----ATT 287
 OY 188 TMLENVK-KCKNPLAMIKLSCGSSQSPMGONVKKVEOLLDAKIEEFTFRKLYVELK 246
 DB 288 TQCGNTKCKKFLANLEL--STREPKVEKNVTLQIELVANVEDEECDRERILN 345
 OY 247 SSPQPHLVFLKKSVALROL-----LPNSQSFIOQ--- 276
 DB 346 ASPQCLIGFLKSLPRLQALYRKELVIEIKPPQHVGLAGLSQDLPRIQAOIRIG 405
 OY 277 -----QCVQOISSDWIATCTTTTSPVTTTSSSQSEKSTIVSGATAPRTVSV---VOT 329
 DB 406 PSQTTTIGQITQVRMI--TPNALGTPRPTIGTTTISKOPPN---IRLPAPRLVTVGRT 460
 OY 330 LNPAGPVGAKAGVTLHSVGPTAA 354
 DB 461 QIP-SLOYPGQANIVQIR--GPQHA 482

```

RESULT 4
HC1_MESAU STANDARD; PRT; 2090 AA.
ID HC1_MESAU
AC P51611;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Host cell factor C1 (HCF) (VP16 accessory protein) (HC1) (VCAF)
(GFP).
GN HCF1.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Chordata; Craniata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
NCBI_TaxId=10036;
RN 111
RP SEQUENCE FROM N.A.
RA Goto H., Nishitani H., Umene K.I., Nakabeppu Y., Nishimoto T.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: UPON LYTIC INFECTION OF PERMISSIVE CELLS, THE HSV
CC TRANSCRIPTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF
CC ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING
CC PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR
CC ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES (BY
CC SIMILARITY).
CC -1- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM
CC 110 TO 150 KDA AND A MINOR 300 KDA POLYPEPTIDE. THE MAJORITY OF N-
CC AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBEIT
CC NONCOVALENTLY, ASSOCIATED (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE
CC SIGNAL (BY SIMILARITY).
CC -1- PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH
CC CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS
CC AT A DEFINED SITE, PCPE/THET, WITHIN THE HCF REPEAT (BY
CC SIMILARITY).
CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC CC EMBL: D45419; BAA08258.1;
CC DR InterPro: IPR003961; FN_III.
CC DR InterPro: IPR001798; Kelch.
CC DR Pfam: PF00041; fn3; 2.
CC DR Pfam: PF01344; Kelch; 5.
CC DR SMART: SM00060; FN3; 2.
CC DR Nuclear protein; Repeat.
CC FT REPEAT 44 89
CC FT REPEAT 93 140
CC FT REPEAT 148 194
CC FT REPEAT 217 265
CC FT REPEAT 266 313
CC FT REPEAT 313 366
CC FT DOMAIN 1010 1448
CC FT REPEAT 1010 1035
CC FT REPEAT 1072 1097
CC FT REPEAT 1101 1126
CC FT REPEAT 1157 1182
CC FT REPEAT 1295 1320
CC FT REPEAT 1320 1348
CC FT REPEAT 1358 1383
CC FT REPEAT 1423 1448
CC FT REPEAT 1423 1448
CC SEQUENCE 2090 AA; 214942 MW; E495E8B1F285E17 CRC64;
Query Match 10.0%; Score 177.5; DB 1; Length 2090;
Best Local Similarity 23.6%; Pred. No. 0.0031;
Matches 106; Conservative 60; Mismatches 144; Indels 139; Gaps 22;

```

```

OY 10 VSAPKVSQGPRLPAPOIVA VKAPNTTIOFPANLQLPCTVLIKNSGPMVSPQOTV 69
DB 537 IGSNPMQSGMAALAAAAATOKIIPSSA---PVLVSVPACTTIVKT-----VAVTPQTT 588
OY 70 TRAEITNSIRPAPVNPOT-----VKICVPRSSQ---LKKVAVTPYKRL 115
DB 589 LPA--TVKVASPVMVSNPRLRLKTAQAQVTSVSSAANTSTRPITTVKSGTVYAAQ 646
OY 116 AQIGTVV-----TVVPRP-----SSQSVAVPTS-----140
DB 647 AQVTVTVGVGVTITVLKSPISVPGSALISNLGKMSVVGKRPQVTSVTAQASIGPV 706
OY 141 -----VTVTPGKPLNTVTL-----KPSLGAASST-----PSNE 170
DB 707 TQIIQTGPPLPAGTILKLVTSADGKPTTITVTAQSGAGSKPILIGISSVSPSTRKPTGT 766
OY 171 PNKAEKSAVQINLSPTMLENKKNFLAML--IKACGSGSPENGVAKRLVQLD 229
DB 767 TITKTPMSALITQAGATGVTSPGKSPITITTKVMTSGTAP-----AKITAV-- 818
OY 230 AKIEAEFTKRL--YVELKSP--QP---HLVPELKSVALROLPLNSQSFIOQCVQOTS 283
DB 819 PKATGCGQGVQVYVVKAGPQGALRTVP-----MSGVRLVTVYVSAAVPAV---- 869
OY 284 SDNVATCTTTVTTSVYVTTTSSQSEKSIIVSGATAPRTVSQVTLNPLAGVGAKGY 343
DB 870 -TTLVVKGTGVTVTLGVTGVTST-----LAGAGHSTSA-----SLATPITLGT 916
OY 344 VTLHS--VGPTA-----ATGTT 359
DB 917 ATLSQVINPTALTVSAAQTTLTAAGLT 945
-----
RESULT 5
HC1_HUMAN STANDARD; PRT; 2035 AA.
ID HC1_HUMAN
AC P51610;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Host cell factor C1 (HCF) (VP16 accessory protein) (HC1) (VCAF)
(GFP).
GN HCF1 OR HCF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxId=9606;
RN 111
RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS), AND PARTIAL SEQUENCE.
RA Wilson A.C., Lamarco K., Peterson M.G., Herr W.;
RT "The VP16 accessory protein HCF is a family of polypeptides processed
RT from a large precursor protein.";
RL Cell 74:115-125(1993).
RN 121
RP SEQUENCE OF 65-2035 FROM N.A. (LONG AND SHORT ISOFORMS).
RC TISSUE-Petal brain;
RA MEDLINE-95130107; PubMed-7829097;
RA Fratini A., Faranda S., Redolfi E., Zucchi I., Villa A.,
RA Patrosso M.C., Strina D., Susani L., Vezzoni P.;
RT "Genomic organization of the human VP16 accessory protein, a
RT housekeeping gene (HCF1) mapping to Xq28.";
RL Genomics 23:30-35(1994).
RN 131
RP PROCESSING.
RA MEDLINE-96033796; PubMed-7590226;
RA Wilson A.C., Peterson M.G., Herr W.;
RT "The HCF repeat is an unusual proteolytic cleavage signal.";
RL Genes Dev. 9:2445-2458(1995).
CC -1- FUNCTION: UPON LYTIC INFECTION OF PERMISSIVE CELLS, THE HSV
CC TRANSCRIPTIVATOR PROTEIN VP16 ASSOCIATES WITH THE OCTAMER MOTIF-BINDING
CC ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING

```

CC PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR
 CC ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES.
 CC -1- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM
 CC 110 TO 150 KDA AND A MINOR 300 KDA POLYPEPTIDE. THE MAJORITY OF N-
 CC AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBEIT
 CC NONCOVALENTLY, ASSOCIATED.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING. IN THE
 CC SHORTER FORM THE N- AND THE C-TERMINAL FRAGMENTS FAIL TO
 CC ASSOCIATE.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN FETAL TISSUES AND THE
 CC ADULT KIDNEY. PRESENT IN ALL TISSUES TESTED.
 CC -1- DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE
 CC SIGNAL.
 CC -1- PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH
 CC CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS
 CC AT A DEFINED SITE, PCE/THER, WITHIN THE HCF REPEAT.
 CC -1- PTM: GLYCOSYLATED. CONTAINS O-LINKED N-ACETYLGLUCOSAMINE RESIDUES.
 CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L20010; NOT_ANNOTATED_CDS.
 CC EMBL: X79198; CNA5790.1; ALT_INIT.
 CC Genew: HGNC:4839; HCFCL1.
 CC MIM: 300019; -
 CC InterPro: IPR003961; FN.III.
 CC InterPro: IPR001798; Kelch.
 CC Pfam: PF00041; fn3.1.
 CC Pfam: PF01344; Kelch; 5.
 CC SMART: SM00060; FN3; 1.
 CC K W Nuclear protein; Repeat; Alternative splicing; Glycoprotein.
 CC FT REPEAT 44 89 KELCH 1.
 CC FT REPEAT 93 140 KELCH 2.
 CC FT REPEAT 148 194 KELCH 3.
 CC FT REPEAT 217 265 KELCH 4.
 CC FT REPEAT 266 313 KELCH 5.
 CC FT DOMAIN 1010 1439 8 X 26 AA APPROXIMATE REPEATS.
 CC FT REPEAT 1010 1035 HCF REPEAT 1.
 CC FT REPEAT 1072 1097 HCF REPEAT 2.
 CC FT REPEAT 1101 1126 HCF REPEAT 3.
 CC FT REPEAT 1158 1183 HCF REPEAT 4 (DEGENERATE).
 CC FT REPEAT 1286 1311 HCF REPEAT 5.
 CC FT REPEAT 1314 1339 HCF REPEAT 6.
 CC FT REPEAT 1349 1374 HCF REPEAT 7 (DEGENERATE).
 CC FT REPEAT 1414 1439 HCF REPEAT 8.
 CC FT VARSPLIC 382 450 MISSING (IN SHORT ISOFORM).
 CC FT CONFLICT 364 564 R -> A (IN REF. 1).
 CC FT CONFLICT 603 603 S -> SVS (IN REF. 2).
 CC FT CONFLICT 1164 1164 P -> S (IN REF. 1).
 CC FT CONFLICT 1873 1873 A -> G (IN REF. 1).
 CC SQ SEQUENCE 2035 AA; 208841 MW; 99207FBB875204C0 CRC64;

Query Match 9.98; Score 175; DB 1; Length 2035;
 Best Local Similarity 24.24; Pred. No. 0.0043;
 Matches 108; Conservative 57; Mismatches 148; Indels 134; Gaps 22;

QY 10 VSAPKVSRRPAPRIVAKAPNTTIOEPANLOLPQGVILKNSGRLMLVSPQOTV 69
 Db 537 IGSSPQSSGMAALAAAAAATOKIPSSR---PYVLSPVAGTTIVKT-----MAVTPGTTT 588
 QY 70 TRAEFTSNITSRPVAPNPOT-----VKICTVPNSSO---LIKRVAVTPVKL 115
 Db 589 LPA--TVKVASPPVMSNPATRMKLTAAAOVGTSSSATNTSTRPIITVHKSGIVTAQO 646
 QY 116 AQIGTIV-----TVPRP-----SSVQSVAVPTS----- 140

Db 647 AQVNTTVGGVTKTITLVKSPISVPGSALISNLKRWVSVOOTRPOVTSVATGASTGPV 706
 QY 141 -----VVITPGKPLNVT-----LKPSISGASST-----PSNE 170
 Db 707 TQIQTGPLPAGTILKLVTSADSKPTTIITTOASGAGTKPILGLISSVSPSTTKPFTT 766
 QY 171 PNLKENSAAVOINLSPMLENVKCKKNFLAML-TKLACSGSQSPENGQNYKRLVEDLLD 229
 Db 767 TIITIMSAIITGAGATGVSSPGISPLITITTKWTSQTAP-----AKITAV-- 818
 QY 230 AKIAEEETKRL-YVELKSP-QP---HLVPLKSVVALROLPLNSQFTQOCVOQTS 283
 Db 819 PKIATGHGOQGVTOVYVKGAPGPGTILRTVP-----MGGRVLTVPYVSAVPRV---- 869
 QY 284 SDVIANCTPTVTSPTVTTVSSQS-----EKSIYSGATPAPTVSVOTLN 332
 Db 870 -TTLVKGITGVTLGLTGTSTSLAGAGHSSTASLAPITTLGTIA--TLSSOYINP 926
 QY 333 LAGPVGAKAGVTLHSVGPATAGTT 359
 Db 927 TA--ITVSAQTTL-----TAAGGLTT 946

RESULT 6
 MUC2_HUMAN
 ID MUC2_HUMAN STANDARD; PRT; 5179 AA.
 AC 002817; 014878;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Mucin 2 precursor (Intestinal mucin 2).
 GN MUC2 OR SMUC.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Intestine;
 RX MEDLINE=94132002; PubMed=8300571;
 RA Gum J.R., Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
 RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
 RT Identification of the amino terminus and overall sequence similarity
 RT to prepro-von Willebrand factor.";
 RT J. Biol. Chem. 269:2440-2446(1994).
 RN [2]
 RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
 RC TISSUE=Colon;
 RX MEDLINE=93016075; PubMed=1400449;
 RA Gum J.R., Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
 RA Kim Y.S.;
 RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
 RT both upstream and downstream of its central repetitive region.";
 RT J. Biol. Chem. 267:21375-21383(1992).
 RN [3]
 RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
 RX MEDLINE=91358717; PubMed=1885763;
 RA Toribara N.W., Gum J.R., Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
 RA Petersen G.M., Kim Y.S.;
 RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
 RT and polymorphism.";
 RT J. Clin. Invest. 88:1005-1013(1991).
 CC -1- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
 CC OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A
 CC PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS
 CC AGENTS AT MUCOSAL SURFACES.
 CC -1- SUBUNIT: MULTIMERIC.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
 CC BRONCHUS, CERVIX AND GALL BLADDER.
 CC -1- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
 CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).

DT 01-FEB-1995 (Rel. 31, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Glucosylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase) (1,4-alpha-D-glucan glucosylase).
 GN STAL OR STA2 OR MAL5 OR YIR019C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 ON NCBI_TaxID=4932;
 RX MEDLINE=67194600; PubMed=106330;
 RA Yamashita I., Nakamura M., Fukui S.;
 RT "Gene fusion is a possible mechanism underlying the evolution of STAL";
 RL J. Bacteriol. 169:2142-2149(1987).
 RN [3]
 RP SEQUENCE OF 1-31 FROM N.A.
 RC STRAIN=SPX101-1C;
 RX MEDLINE=89031230; PubMed=3141213;
 RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
 RT "Similar short elements in the 5' regions of the STA2 and SOA genes from Saccharomyces cerevisiae";
 RL FEBS Lett. 239:179-184(1988).
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-glucose residues successively from non-reducing ends of the chains with release of beta-D-glucose.
 CC -1- SIMILARITY: TO S.POMBE SPBC215.13.
 CC -1- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z38061; CA86176.1; -
 DR EMBL: M16164; AAA35014.1; -
 DR EMBL: M16165; AAA35015.1; -
 DR EMBL: X13857; CAA32069.1; -
 DR PIR: B26877; B26877.
 DR PIR: A26877; A26877.
 DR PIR: S48478; S48478.
 DR SDD: S0001458; MUC1.
 KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein; Signal; Multigene family.
 KM SIGNAL: Multigene family.
 FT CHAIN 1 21 POTENTIAL.
 FT DOMAIN 22 1367 GLUCOMYLAZE S1/S2.
 FT CARBOHYD 210 1367 SER/THR-RICH.
 FT CARBOHYD 817 817 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61AA9D CRC64;
 Query Match 9.18; Score 162; DB 1; Length 1367;
 Best Local Similarity 21.28; Pred. No. 0.016;
 Matches 87; Conservative 62; Mismatches 181; Indels 80; Gaps 13;
 QY 2 TLVTKAVPVSAPKVSQGRLPAP-----QIVAVKAPNTTQIOFANLQLPCTVLIK 54
 Db 618 TTESSAPVSTSTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTE 677

QY 55 SNSGFLMLVSPQOT--VTRAETTSNITSRPVAPNPQIVICVTPNNSOLIKRVAV-- 109
 Db 678 SSASAPVSTSTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPV 737
 QY 110 -TPVKKLAQIGTVYTT-----VPKSSVQS-----YAVPTSVYVTPGKPLMT- 152
 Db 738 PTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTP 797
 QY 153 -----VTLKPSILGASPTSPNEPNLKAENSAVOINLSPMLAEVKKCKNPLA 201
 Db 798 SSSTTESSAPVPTPSSSNITSSAPSTPSSSTTESSSVY---PTSSSTTESS--- 851
 QY 202 MLKILACSGSOS-----PENGQVKKLVQLLDKATIEAEETFKLYELKSSPOPLV 254
 Db 852 --APVSSSTTESSAPVPTPSSSNITSSAPSIIPSSSTTESFSGTVTPSSSK----- 904
 QY 255 PFLKRSVALKQLPNQSGFIOQCQOQSSDMVATCTTYTTSR---VTTTVSSSGSE 311
 Db 905 -----YPGSOT--ETVSSSTETTYVPTKTTTSVTPSTTTTTCVCSGTGN 949
 QY 312 KSIIVSGATAPRTV--SVQTLNPLAGPVGAKAGVYTLHSGPTAATGCTT 359
 Db 950 SAGETTSCTSKRTVYTTVPTTTTTSVTTSSTTTTTCVCSGTGNAGETT 999
 RESULT 8
 ID YS89 CAEL STANDARD PRT: 3178 AA.
 AC Q09624; Q09625; Q09624;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUN-2002 (Rel. 41, last sequence update)
 DT 15-JUN-2002 (Rel. 41, last annotation update)
 DE Hypothetical protein ZK945.9 in chromosome II.
 GN ZK945.9/ZK945.10.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 CC Rhabditidae; Peloderinae; Caenorhabditis.
 ON NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Wilkinson-Sproat J.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP REVISIONS.
 RA Durbin R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: CONTAINS 1 GPS DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z48544; CAB70192.1; -
 DR EMBL: Z48582; CAB70192.1; JOINED.
 DR EMBL: Z48582; CAB70201.1; -
 DR EMBL: Z48544; CAB70201.1; JOINED.
 DR WormPep: ZK945.9; CE25697.
 DR InterPro: IPR002111; Cat_channel_Tyrl.
 DR InterPro: IPR001024; Lipoxigenase_LH2.
 DR InterPro: IPR000636; M-channel_nlg.
 DR InterPro: IPR000203; PKD_cys_rich.
 DR Pfam: PF00520; Ion_trans_1.
 DR Pfam: PF01477; PLAT_1.
 DR Pfam: PF01825; GPS_1.
 DR SMART: SM00303; GPS_1.
 DR SMART: SM00308; LH2_1.
 KW Hypothetical protein; Transmembrane.

FT DOMAIN 266 1196 SER/THR-RICH.
FT DOMAIN 1105 1241 GLY/SER-RICH.
FT DOMAIN 2071 2120 GPs.
FT TRANSMEM 51 30 POTENTIAL.
FT TRANSMEM 2139 73 POTENTIAL.
FT TRANSMEM 2348 2161 POTENTIAL.
FT TRANSMEM 2348 2367 POTENTIAL.
FT TRANSMEM 2390 2412 POTENTIAL.
FT TRANSMEM 2451 2468 POTENTIAL.
FT TRANSMEM 2483 2505 POTENTIAL.
FT TRANSMEM 2567 2589 POTENTIAL.
FT TRANSMEM 2836 2858 POTENTIAL.
FT TRANSMEM 2939 2961 POTENTIAL.
FT TRANSMEM 2976 2998 POTENTIAL.
FT TRANSMEM 3038 3060 POTENTIAL.
SQ SEQUENCE 3178 AA; 344726 MW; 88239436D03666CD CRC64;

Query Match 8.9%; Score 157; DB 1; Length 3178;
Best Local Similarity 20.5%; Pred. No. 0.091;

Matches 84; Conservative 64; Mismatches 188; Indels 74; Gaps 13;

OY 2 TLVTKAPVAPKPVSSGRLPAPQIVAKAPNTTIOFANLQLPFGIVLKSNSGPLM 61
DB 356 TLSTSTPTTPTTPTTSTLSSLPNAICSYLDETTSTTTMLTSTTEPSTTTE 415
OY 62 LVSPOQVTRAEFTSNITSRPAVPA--NPQVKTCTVPNSSOLIKKAVTPVKLAQIG 119
DB 416 VITSTSTVTTEPTTTLTSTASTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 465
OY 120 TVVTVTPKRVSSVQAVAV-PTSVVTVTPKPLVTTLKPSLGASSTPSENPNAENS 178
DB 466 SSSSTVTPTST 521
OY 179 A-AVOINLSPTMLENKKCNFLAMLIKACSGSOSPEMGNKKLVEQLDA----- 230
DB 522 ASSTGSTSTQSSSTTKSET-----TTSSDGTNDF-YFEKATTTFTYDSTVNLTL 573
OY 231 -----KIAEEFTKLYE-----LKSSPQPHL-----VFLKKSVALR 265
DB 574 NSGLIGIGVOTSIETCSPTSSNNVSTTKGACFTKSVSMRLGCTYPASTFVGNTFR 633
OY 266 QLLPNS-----QSFIOQCVOQTSSMVIATCTTIVTSP---VTTTVSSSQ 309
DB 634 ATTTTDDKKVYVYVAVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 693
OY 310 SEKSIIIVGATAPRVSVQTLNPLAGPVGAKAVVLHSGPTAAAGCTT 359
DB 694 STRSDSTTTSAGSTTTLQESTTSESTTDSSTTTTISDTSTSSPSST 743

RESULT 9

VG50_HSV1L
ID VG50_HSV1L STRAND: PRT; 670 AA.
AC 000130;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical gene 50 protein.
GN 50
OS Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Ictalurid Herpes-like viruses.
OX NCB1_TaxID=10401;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Auburn 1;
RX MEDLINE=92087490; PubMed=1727613;
RA Davison A.J.;
RT "Channel catfish virus: a new type of herpesvirus.";
RL Virology 186:9-14(1992).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M75136; AAA88153.1; -

DR PIR: F36791; F36791.

KW Hypothetical protein; Repeat.

FT REPEAT 143 158
FT REPEAT 171 186
FT REPEAT 200 214
FT REPEAT 215 233
FT REPEAT 234 252
FT REPEAT 253 268
FT REPEAT 279 293
FT REPEAT 294 309
FT REPEAT 320 334
FT REPEAT 335 349
FT REPEAT 362 376
FT REPEAT 377 391
FT REPEAT 392 406
FT REPEAT 407 421
FT REPEAT 422 436
FT REPEAT 437 452
FT REPEAT 464 477
FT REPEAT 478 493
FT REPEAT 504 517
FT REPEAT 518 531
FT REPEAT 532 545
FT REPEAT 546 559
FT REPEAT 560 573
FT REPEAT 574 587
FT REPEAT 588 601
FT REPEAT 602 615
FT REPEAT 616 629
SQ SEQUENCE 670 AA; 64174 MW; 2B64A781C519E8B4 CRC64;

Query Match 8.8%; Score 155.5; DB 1; Length 670;
Best Local Similarity 24.5%; Pred. No. 0.017;

Matches 91; Conservative 33; Mismatches 129; Indels 119; Gaps 17;

OY 16 VSSGRLPAPQIVAKAPNTT-----IOPANLQLPFGIVLKSNSGPILVSPQOTV 70
DB 248 VTTTPAMPAG---ANDANTTTATPTGANDTANTVMPAGATDVVTTTPAM-----PT 297
OY 71 RAFTSNITSRPAVPANPQTVKICTVPNSSOLIKKAVTP-----VKLAQIG 119
DB 298 GANDTANIT--ATPAGANDTANTVMPAGATDV--VTTTPAMPAGANDTANTKPA GST 353
OY 120 TVVTVTPKRVSSVQAVAVPT---SVYTVPKRP---LNTVTLKRS-SLGASSTPSNEP 171
DB 354 DTVVTTTP-----AMPGATDVVTTTPAMPTGATDVVTTTPAMPPTGATDVVTTT 405
OY 172 NLKAENSAVOINLSPTMLENKKCNFLAMLIKACSGSOSPEMGNKKLVEQLDDAK 231
DB 406 PAKPAGANGVTTTPAM-----PAGAN----- 428
OY 232 IEAEFTKRLKLYELKSSPQPHLVFLKKSVALROLPLNSQSFIOQCVOQT-SSDNVIAT 290
DB 429 -----DIVVTTAPATPAGANDTANTVTKPTGATDVVTT 461
OY 291 CTT--TVTSPVYTVTVSSQSEKSIIVSGAT-APRVSVQTLNPLAGPVGAKAVGVTTLH 347
DB 462 ATVPKPGATGTVTTTAKPAGANDTANTVTKPTGATGTVTTT---AKPTGA-TGVTVA 516
OY 348 SVGPATAGGTT 359
DB 517 TAKPTGATGTVT 528

RESULT 10
VG03_EBVAB

ID VGP3_EBVA8 STANDARD: PRT: 866 AA.

AC 007284;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DE Envelope glycoprotein GP340 (Membrane antigen) (MA).

GN BLF1.

OS Epstein-Barr virus (strain AG876) (Human herpesvirus 4), and

OS Epstein-Barr virus (strain P3HR-1) (Human herpesvirus 4).

OC Viruses; dsDNA viruses, no RNA stage: Herpesviridae;

OC Gammaherpesvirinae: Lymphocryptovirus.

OX NCBI_TaxID=82830, 82829;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-AG876, and P3HR-1;

RX MEDLINE=93331716; PubMed=8393237;

RA Lees J.F., Arrand J.E., Pepper S.V., Stewart J.P., Mackett M.,

RA Arrand J.R.;

RT "The Epstein-Barr virus candidate vaccine antigen gp340/220 is highly

RT conserved between virus types A and B.";

RL Virology 195:578-586(1993).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-P3HR-1;

RA Klein K., Mueller-Lantsch N.;

RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CR2 RECEPTOR ON HUMAN

CC B-CELLS.

CC -1- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL

CC ENVELOPE.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

DR EMBL: L07922; AAA02783.1; -

DR EMBL: L07923; AAA02787.1; -

DR EMBL: X67776; CAA47986.1; -

KM Membrane; Glycoprotein; Antigen; Late protein.

FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 169 169 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 195 195 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 229 229 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 328 328 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 386 386 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 435 435 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 443 443 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 533 533 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 589 589 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 624 624 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 680 680 N-LINKED (GLCNAC. . .) (POTENTIAL).

SO SEQUENCE 886 AA; 92388 MW; 4394F6130DECCABA CRC64;

Query Match 8.7%; Score 154.5; DB 1; Length 886;

Best Local Similarity 23.6%; Pmed. No. 0.027;

Matches 96; Conservative 42; Mismatches 162; Indels 107; Gaps 19;

QY 2 TLVTKVAPVSAAPKVVSGPRRLPAPQIVAKAINTT-----IQFPANQLPQGVILIKSN 56

DB 414 TTHKVVIRSKAPESSTTSPTL---NTGFAAPNTTGLPSSHTVPTNLTPAST----- 464

QY 57 SGPLMLVSPQOVTVAETTSNITSRAVPA-----NPQYKICTVPNS 100

DB 465 -GPTVSTADVTSPTPAGTSSG--ASPVPSPSPRDNESKAPDMTSPTSATVTPTPNAT 521

QY 101 SOLIKKAVTPPKKLAQIGTVVT---TVPKP---SSVQSAVAPTSVTV--TPGR--PLN 151

DB 522 SPTPAVTPTPPNATSPTKTSPTSVAVTTPPNATSPPAVTPPNATIPTKGTSPTS 581

QY 152 TVTTLKPS-----SLGASSTPSNEPMLKAENSAVOINTSP-----TMLENYKKCKN 198

DB 582 AVTTTPPNATSPVETSTPQANTTHTLGGTSSTVTVSPKRNATSAVTTGQHNITSSST 641

QY 199 FLAMLIKACSGSGSPENGQVKLVLEQLDAKIEEFTKLYELKSSPOPILVPLK 258

DB 642 SSMILRPSISSETLSPTSNDSTS-----HMLLTSAHP----- 675

QY 259 KSVVALRQLLRNSQSFIOQCVOQTSDDAVIATCTTTVTSPTVTTVS-----SSQSEK-- 312

DB 676 -----TGENITGVTPASTSTHVV---STSSPAPRGTTSQASGPNSSSTKRG 722

QY 313 SIIVGATAPRTVSQVTLNPLAGPVGAKAGVYTLHSVPTA--ATGG 357

DB 723 EVNVTKGTPPKNAT-----SPQA-PSGQKTAVPVTVSTGKANSTTGG 764

RESULT 11

SP2_HUMAN STANDARD: PRT: 606 AA.

ID SP2_HUMAN

AC 002086;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Transcription factor Sp2.

GN SP2 OR KIA00048.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Bone marrow;

RX MEDLINE=96051396; PubMed=7584044;

RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,

RA Seki N., Kawarabayashi Y., Ishikawa K.-I., Tabata S.;

RT "Prediction of the coding sequences of unidentified human genes. II.

RT The coding sequences of 40 new genes (KIA00041-KIA00080) deduced by

RT analysis of cDNA clones from human cell line KG-1.";

RL DNA Res. 1:223-229(1994).

RN [2]

RP SEQUENCE OF 112-606 FROM N.A.

RX MEDLINE=93024366; PubMed=1341900;

RA Kingsley C., Winoto A.;

RT "Cloning of c-fos-binding proteins: a novel Sp1 multigene family

RT regulating T-cell receptor gene expression.";

RT Mol. Cell. Biol. 12:4251-4261(1992).

CC -1- FUNCTION: BINDS TO GC BOX PROMOTERS ELEMENTS AND SELECTIVELY

CC ACTIVATES MRNA SYNTHESIS FROM GENES THAT CONTAIN FUNCTIONAL

CC RECOGNITION SITES.

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER

CC PROTEINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@slb.ch.

CC EMBL: D28588; BAA05923.1; -
 CC EMBL: M97190; AAA3629.1; -
 DR PIR: A44489; A44489.
 DR HSSP: P08047; 1SP2.
 DR TRANSFAC: T02356; -
 DR Genew: HGNC:11207; SP2.
 DR MIM: 601801; -
 DR InterPro: IP000822; Znf_C2H2.
 DR Pfam: PF00096; Zf-C2H2; 3.
 DR PRINTS: PR00048; ZINCINGER.
 DR ProDom: PD000003; Znf_C2H2; 2.
 DR SMART: SM00355; Znf_C2H2; 3.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 3.
 DR Transcription regulation: Activator; zinc-finger; Metal-binding;
 KW DNA-binding; Nuclear Protein; Repeat.
 FT DOMAIN 518 600 ZINC_FINGERS.
 FT ZN_FING 518 542 C2H2-TYPE.
 FT ZN_FING 548 572 C2H2-TYPE.
 FT ZN_FING 578 600 C2H2-TYPE.
 SQ SEQUENCE 606 AA; 64153 MW; A27C6D460D36E186 CRC64;

Query Match Best Local Similarity 24.9%; Pred. No. 0.021; Score 153; DB 1; Length 606;
 Matches 96; Conservative 52; Mismatches 137; Indels 100; Gaps 19;

OY 7 VAVPVSAPVSSGPRLLP-----DQIVAVKAPNTTIOFPANL-----OLPGGVLIKS 55
 DB 104 VFAIQPTMIKNGTRSNANIQVAPQI---QASNOTIQVQNLNQIIRGT----- 155
 OY 56 NSGFLMLVSPQQTIVTAETTSNTTSRPANPQIVKICIVPSSSOLLKKAIVTPYK-- 113
 DB 156 -----NQAITTPSPSHKPVPIKPAPIO-----KST-----TTTPVQSG 190
 OY 114 ----KLAQIGTVVTVTPKPSVOS--VAVPTSVTVTPKPLN--TYTTLKPSLIGASST 166
 DB 191 ANVVKILTGCGGANTLPLPANNVNASDTGAPQIOLTESPTPLSKTKKAKKSLPASOP 250
 OY 167 PSNEPLKAENSAVOINLSPWLENVKKCKNEFLMLIKLACSGSSPEMGQVKKLVQ 226
 DB 251 P-----VAAVEQVETVLETTADNIIQGNLLIV-----QSPGGGQAVVQVQ 295
 OY 227 LLDAKTEAEFFRKLYVELKSSPQPHLVPLKSSVALNQLDPN-----SQSTIQCVQ 261
 DB 296 VVPPKAEQOQVQO-----IPQAL-----RVQVAAASATLPTVPQKPSQNPQIOAAEP 342
 OY 282 TSSDVIAT-----CTTVTTSFVVT-----TVSSSQSEKSIIVGATAPRTVSV--QTL 330
 DB 343 TPTQVYIRTPSGGVQVIVQDSPRAAATNSNTTCSPPASRAPHLSGTSKKHSAILRKE 402
 OY 331 NPLAGPVGAKAG-VVTLHSVGPTAA 354
 DB 403 RPL--PKIAPAGSIISLNAQLAA 425

RESULT 12
 MTG8_HUMAN STANDARD; PRT; 604 AA.
 AC Q06455; Q06455; Q16239; Q92479; Q16346; Q16347; Q14873; Q9BR20;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT MTG8 protein (ETO protein) (Eligth twenty one protein) (Cyclin D
 DE related protein).
 GN CBFA2T1 OR MTG8 OR ETO OR CDR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANT AML1-MTG8/ETO FUSION IN AML-M2.

RC TISSUE=retal brain;
 RX MEDLINE=93327761; PubMed=8334990;
 RA Miyoshi H., Kozu T., Shimizu K., Enomoto K., Maseki N., Kaneko Y.,
 RA Kamada N., Ohki M.;
 RT "The t(8;21) translocation in acute myeloid leukemia results in
 RT production of an AML1-MTG8 fusion transcript.";
 RL EMBO J. 12:2715-2721(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98326315; PubMed=9661669;
 RA Wolford J.K., Prochazka M.;
 RT "Structure and expression of the human MTG8/ETO gene.";
 RL Gene 212:103-109(1998).
 RN [3]
 RP SEQUENCE OF 31-604 FROM N.A.
 RX MEDLINE=94185017; PubMed=8137293;
 RA Erickson P.F., Robinson M., Owens G., Drabkin H.A.;
 RT "The ETO portion of acute myeloid leukemia t(8;21) fusion transcript
 RT encodes a highly evolutionarily conserved putative transcription
 RT factor.";
 RL Cancer Res. 54:1782-1786(1994).
 RN [4]
 RP SEQUENCE OF 30-604 FROM N.A.
 RC TISSUE=Muscle;
 RA Strausberg R.;
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 334-432 FROM N.A., AND VARIANT AML1-MTG8/ETO FUSION.
 RX MEDLINE=95002916; PubMed=7919324;
 RA Tighe J.E., Calabi F.;
 RT "Alternative, out-of-frame runt/MTG8 transcripts are encoded by the
 RT derivative (8) chromosome in the t(8;21) of acute myeloid leukemia
 RT M2.";
 RL Blood 84:2115-2121(1994).
 RN [6]
 RP VARIANT AML1-MTG8/ETO FUSION IN AML-M2.
 RX MEDLINE=95329434; PubMed=7541640;
 RA Eira T., Asou N., Kunitada T., Yamasaki H., Asou H., Kamada N.,
 RA Nishikawa S.-I., Yamaguchi K., Takatsuki K.;
 RT "Identification of two transcripts of AML1/ETO-fused gene in t(8;21)
 RT leukemic cells and expression of wild-type ETO gene in hematopoietic
 RT cells.";
 RL Genes Chromosomes Cancer 13:25-33(1995).
 RN [7]
 RP VARIANT AML1-MTG8/ETO FUSION IN AML-M2.
 RX MEDLINE=93357484; PubMed=8353289;
 RA Kozu T., Miyoshi H., Shimizu K., Maseki N., Kaneko Y., Asou H.,
 RA Kamada N., Ohki M.;
 RT "junctions of the AML1/MTG8(ETO) fusion are constant in t(8;21) acute
 RT myeloid leukemia detected by reverse transcription polymerase chain
 RT reaction.";
 RL Blood 82:1270-1276(1993).
 RN [8]
 RP VARIANT AML1-MTG8/ETO FUSION IN AML-M2.
 RX MEDLINE=93046086; PubMed=1423235;
 RA Nilsson P.E., Watkins P.C., Sacchi N.;
 RT "Transcriptionally active chimeric gene derived from the fusion of
 RT the AML1 gene and a novel gene on chromosome 8 in t(8;21) leukemic
 RT cells.";
 RL Cancer Genet. Cytogenet. 63:81-88(1992).
 CC -1- FUNCTION: PUTATIVE TRANSCRIPTION FACTOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: MTG8A AND MTG8B (SHOWN HERE);
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: MOST ABUNDANTLY EXPRESSED IN BRAIN. LOWER
 CC LEVELS IN LUNG, HEART, TESTIS AND OVARY.
 CC -1- DISEASE: INVOLVED IN ACUTE MYELOID LEUKEMIA (AML-M2) BY A
 CC CHROMOSOMAL TRANSLOCATION t(8;21)(Q26;Q22) THAT INVOLVES CBFA2T1
 CC AND PROTEIN AML1.
 CC -1- SIMILARITY: CONTAINS 1 MYND-TYPE ZINC FINGER.
 CC -1- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Hematol.;
 CC WWW="http://www.infobogen.fr/services/chromacancer/Genes/ETO.html".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (see [http://www.isb-sib.ch/announcement/](http://www.isb-sib.ch/announcement/CC)
CC or send an email to license@isb-sib.ch).

FT	SEQUENCE	604 AA:	67566 MW:	C3D2452F96E5679 CRC64:
DR	EMBL; D14821; BAA03558.1; ALT_SEQ.			
DR	EMBL; D13979; BAA03089.1; ALT_SEQ.			
DR	EMBL; D14289; BAA03247.1; ALT_INIT			
DR	EMBL; AF018282; AAC28932.1; -			
DR	EMBL; AF018271; AAC28932.1; JOINED			
DR	EMBL; AF018273; AAC28932.1; JOINED			
DR	EMBL; AF018274; AAC28932.1; JOINED			
DR	EMBL; AF018275; AAC28932.1; JOINED			
DR	EMBL; AF018276; AAC28932.1; JOINED			
DR	EMBL; AF018278; AAC28932.1; JOINED			
DR	EMBL; AF018279; AAC28932.1; JOINED			
DR	EMBL; AF018281; AAC28932.1; JOINED			
DR	EMBL; AF018282; AAC28931.1; -			
DR	EMBL; AF018287; AAC28931.1; JOINED			
DR	EMBL; AF018273; AAC28931.1; JOINED			
DR	EMBL; AF018274; AAC28931.1; JOINED			
DR	EMBL; AF018275; AAC28931.1; JOINED			
DR	EMBL; AF018276; AAC28931.1; JOINED			
DR	EMBL; AF018277; AAC28931.1; JOINED			
DR	EMBL; AF018278; AAC28931.1; JOINED			
DR	EMBL; AF018279; AAC28931.1; JOINED			
DR	EMBL; AF018281; AAC28931.1; JOINED			
DR	EMBL; D43638; BAA07755.1; -			
DR	EMBL; BC005850; AAH05850.1; ALT_INIT			
DR	EMBL; X79990; CA556311.1; -			
DR	EMBL; S74096; AAB32126.1; -			
DR	EMBL; S74092; -; NOT_ANNOTATED_CDS.			
DR	EMBL; S78158; AAB34819.2; ALT_SEQ.			
DR	EMBL; S78159; AAB34820.2; ALT_SEQ.			
DR	EMBL; D14822; BAA03559.1; ALT_SEQ.			
DR	EMBL; D14823; BAA03560.1; ALT_SEQ.			
DR	EMBL; S50186; -; NOT_ANNOTATED_CDS.			
DR	GeneW; HGNC:1535; CBFA2T1.			
DR	MIM; 133435; -			
DR	InterPro; IPR003894; TAF_hom.			
DR	InterPro; IPR002893; ZnF_MYND.			
DR	Pfam; PF01753; Zf-MYND; 1.			
DR	SMART; SMO0349; TAFH; 1.			
KW	Zinc-finger; DNA-binding; Nuclear protein; Alternative splicing; Chromosomal translocation; Proto-oncogene; Transcription regulation.			
FT	DOMAIN	57	60	POLY-PRO.
FT	DOMAIN	102	108	POLY-SER.
FT	DOMAIN	290	297	POLY-PRO.
FT	DOMAIN	408	413	POLY-SER.
FT	SITE	30	31	BREAKPOINT FOR TRANSLLOCATION TO FORM MYD-MTG8 IN AML-M2.
FT	2N_FING	515	551	MYD-TYPE.
FT	VASAPLIC	1	29	MISXRRNALSLVIGCRKKKNFYEQ -> MP (IN ISOTORM MTG8A).
SO	SEQUENCE	604 AA:	67566 MW:	C3D2452F96E5679 CRC64:

```

Oy 204 IKACSGSGSEPMONKVLLEOLDIKIEEEDTRILYAEKSSPPHLYPLFKKSVA 263
      |||::||::||::||::||::||::||::||::||::||::||::||::||::
Db 134 QGF--GNDISPEIGERKTVTLGVLNLTLLIEEHSKLOEETNPPLPRFVIFLKAULPL 191
      |||::||::||::||::||::||::||::||::||::||::||::||::||:
Oy 264 L-ROLLENSQSFQ---QCVQQTSSDMVATICTTYYTTSPPV 301
      |||::||::||::||::||::||::||::||::||::||::||::||::||:
Db 192 LQRELLHCARLAKONPAOYLAKHQLLDSTSPVDSSTELL 233
      |||::||::||::||::||::||::||::||::||::||::||::||::||:

```

```

RESULT 13
YK82_YEAST
ID YK82_YEAST STANDARD; PRT: 1169 AA.
AC P36170;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Hypothetical 122.2 kDa protein in STR1 3' region precursor.
GN YKR102W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Gallion L., Dujon B.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
CC -1- SUBCELLULAR LOCATION: CELL WALL (BY SIMILARITY) .
CC -1- PPM: EXTENSIVELY O-GLYCOSYLATED (BY SIMILARITY) .
CC -1- SIMILARITY: STRONG, TO YEAST PROTEIN FOL1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

[illegible]

Db 801 SSSSESHRTPTTSSSEGIKSSGVIEBTSSTSSSFHHTSTASTSVQTSQFVTPSP 860
 QY 150 LNTVTTLKSSSG-ASSTPSNEPNLKAENSAVOINLSPTMLENVKCKNFLAMLIK 205
 Db 861 ISYVA--FRSTGLNSQSTESTNSKSTMSSENSASV----- 893
 QY 206 LACSSGSPQMGNVKKLVEQLDAKIEAE--FTR--KIYVELKSPQHLVPELK 258
 Db 894 MPSSSATSPKRG-----KVTSDERSSGFSRDRRTTYRMTSETPSTN---EQ 936
 QY 259 KSVVALROLPLPNSQSF1QOCVOQTSSDMVATCTTV-----TT-----SPVTTTVSS 307
 Db 937 TLLITVSSCESNSCS-----NTVSSAVSTTTTITNGITTYTTCPLSATLTLTVSK 989
 QY 308 SOS-EKSIIVSGATAPRVSVOTLNP--LAGPYGAKAGVTTLHSGPTAAT 355
 Db 990 LSEKRTLTITVSCSEGVCSFETASPAIVATATVNDVTVYSTWSPQAT 1040

RESULT 14

CPN_DROME STANDARD; PRT; 865 AA.

ID CPN_DROME 002910;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Calphotin.
 GN CPN OR CAP.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Braachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Canton-S;
 RX MEDLINE=93165729; PubMed=8094559;
 RA Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
 RT "Calphotin: a Drosophila photoreceptor cell calcium-binding protein.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Canton-S;
 RX MEDLINE=93165730; PubMed=8434015;
 RA Ballinger D.G., Xue N., Harshman K.D.;
 RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds
 calcium and contains a leucine zipper.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
 CC -1- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO
 REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL
 OF CA+2 PER MOL OF PROTEIN.
 CC -1- SUBUNIT: HOMODIMER (PROBABLE).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
 CC -1- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
 COMPOUND EYES AND OCELLI.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
 DEVELOPMENT.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: L02111; AAA28405.1;
 DR EMBL: L05080; AAA28420.1;
 DR PIR: A47282; A47282.
 DR Flybase: FBgn010218; Cpn.
 KW Calcium-binding.
 FT CONFLICT 36 36 A -> AVAPAVVA (IN REF. 2).

FT CONFLICT 43 43 I -> T (IN REF. 2).
 FT CONFLICT 64 64 I -> A (IN REF. 2).
 FT CONFLICT 76 76 T -> A (IN REF. 2).
 FT CONFLICT 100 100 P -> PP (IN REF. 2).
 FT CONFLICT 126 127 VO -> AP (IN REF. 2).
 FT CONFLICT 154 154 I -> V (IN REF. 2).
 FT CONFLICT 160 160 S -> T (IN REF. 2).
 FT CONFLICT 534 534 A -> E (IN REF. 2).
 FT CONFLICT 699 699 I -> T (IN REF. 2).
 FT CONFLICT 703 703 V -> L (IN REF. 2).
 FT CONFLICT 721 721 D -> E (IN REF. 2).
 SQ SEQUENCE 865 AA; 84781 MW; 2110417E080E7CFE CRC64;

Query Match 8.5%; Score 150; DB 1; Length 865;
 Best Local Similarity 22.0%; Pred. No. 0.05;
 Matches 90; Conservative 59; Mismatches 176; Indels 84; Gaps 16;

QY 1 GTLVTKV-APVSA-----PKVSSCPRLPAP-----QIYAVKAPNTTIOFANLQ 45
 Db 4 GTLPSPVSAVPAAPVTPSAVAAPOVPSAAPVAPAPAPAPAPAPAPAPAPAPAPAP 63
 QY 46 LPPTGLIKNSGRLMLVSPQQTVAETTSNTTSRPAVPANQTVKICTVPS-SSQL 104
 Db 64 IPAPAPIAASVTPVSAVAP--PYVAAPTP-----PAASPTVPAVVAQIYVAPSAVPA 115
 QY 105 KKAIVTPVKKL-AQIGTTVTVPKRSSVOSVAVPPISVTV-VTPGKPLNTVTTLKSSSG 162
 Db 116 PVAAPATPTPVQIIPVAPVATPTPVASAPTPPAVTPVIVISVTPVPAVPAVPAVPA 175
 QY 163 ASSTPSNEPNLKAENSAVOINLSPTMLE-----NVKKCKNFLAMLIKACS 209
 Db 176 VAAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAV 227
 QY 210 GSQSPKMGVKKLVEQLDAKIEAEFTKRLVLELKSSPQHLVPELKSVVALROLPL 269
 Db 228 ---PEVSVATKPLAAAEHVVAAPATETPPVAPAAASHVSAVPAVETAVVA-----P 278
 QY 270 NSQSF1QOCVOQTSSDMVATCTTVTTSPVTTTVSSQSEKSIIV---SGATAPRV 325
 Db 279 VSAS-----TEPPVAAATLTTAETP-PALAPVAESQVAAVNTVATPPTAPPEPETI 329
 QY 326 S-----VQTLNPLAGPYGAKAGVTTLHSGPTAATGTT 359
 Db 330 APPVVAETPPVSAVNAAEITPPVPPVPAABSIAPVATTPVPAATLAVT 378

RESULT 15

MTG8 MOUSE STANDARD; PRT; 577 AA.

ID MTG8 MOUSE 061909;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE MTG8 protein.
 GN CBFA2T1 OR CBFA2T1H OR MTG8.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ICR; TISSUE-Brain;
 RX MEDLINE=96121389; PubMed=8575770;
 RA Niwa-Kawakita M., Miyoshi H., Gotoh O., Matsushima Y., Nishimura M.,
 RA Shiba H., Ohki M.;
 RT "Cloning and gene mapping of the mouse homologue of the CBFA2T1 gene
 associated with human acute myeloid leukemia.";
 RL Genomics 29:755-759(1995).
 CC -1- FUNCTION: PUTATIVE TRANSCRIPTION FACTOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- SIMILARITY: CONTAINS 1 MYND-TYPE ZINC FINGER.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: D32007; BAA06774.1; -
 DR MGD; MG1:104793; Cbfa2t1h.
 DR InterPro: IPR003894; TAF_hom.
 DR InterPro: IPR002893; Znf_MYND.
 DR Pfam: PF01753; Zf-MYND; 1.
 DR SMART; SM00549; TAFH; 1.
 KW Zinc-finger; DNA-binding; Nuclear protein; Transcription regulation.
 FT DOMAIN 30 33 POLY-PRO.
 FT DOMAIN 75 81 POLY-SER.
 FT DOMAIN 263 270 POLY-PRO.
 FT DOMAIN 380 386 POLY-SER.
 FT ZN_FING 488 524 MYND-TYPE.
 SQ SEQUENCE 577 AA; 64337 MW; 2656F185318C4B11 CRC64;

Query Match

Best Local Similarity 8.3%; Score 147; DB 1; Length 577;
 Matches 65; Conservative 39; Mismatches 90; Indels 28; Gaps 10;

OY 89 QTVKICTVPNNSSQLIKKVAVTPYKLAQIGTIVVTTPKPSYQSVAVPTSVVVTTPGK 148
 Db 4 RTEKSTMPDSDPVVKQSRLLP-----PAMPPEPTTQG-APRTSFTPT--- 47
 OY 149 PLNTVTTLKPSL-GASTP---SNEPNLKAENSAVOINLSPTM-LENVKKCKNFLAML 203
 Db 48 TLNGTSHSPALNGAPSPGFSNGSPSSSSSIANO-QLPPACGARQLSKLRFLLTL 106
 OY 204 IKLACSGSPPEMGONKYLEQLDAKIEAEFTFKLYELKSSPOHIVPFLKSSVYA 263
 Db 107 QQF--GNDISPEIGERYRTLVGLVNSTLTIEEFHSLQEAATNPPLRPVIFPLKANLPL 164
 OY 264 L-ROLPNSSQFIQ--QCVQOTSSDMVIANCTTTVTTSPPV 301
 Db 165 LQRELHCARLAKQNPQVLAQHQLLLDASTSPVDSSELL 206

Search completed: February 16, 2003, 21:56:21
 Job time : 15.9186 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:31:01 ; Search time 25.2146 Seconds

(Without alignments)
2933.659 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_359

Sequence: 1 GTLVTKVAPVSAAPKVS...KAGVTLHSGVPTAATGTT 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: SP_archea:*
- 2: SP_bacteria:*
- 3: SP_fungi:*
- 4: SP_human:*
- 5: SP_invertebrate:*
- 6: SP_mammal:*
- 7: SP_mhc:*
- 8: SP_organelle:*
- 9: SP_phage:*
- 10: SP_plant:*
- 11: SP_rodent:*
- 12: SP_virus:*
- 13: SP_vertebrate:*
- 14: SP_unclassified:*
- 15: SP_virus:*
- 16: SP_bacteriophage:*
- 17: SP_archea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	497	28.0	662	11	091W6 mus musculus
2	194.5	11.0	851	5	08T9E0 drosophila
3	182	10.3	1795	5	076894 drosophila
4	181.5	10.2	1275	5	076602 caenorhabd
5	181	10.2	5374	11	099ND0 mus musculus
6	177.5	10.0	1306	11	09JX31 mus musculus
7	175.5	9.9	709	4	09NVJ9 mus musculus
8	174.5	9.8	1979	11	09OY35 mus musculus
9	173.5	9.8	528	6	029071 sus scrofa
10	172.5	9.7	1349	4	08WQ4 homo sapien
11	171.5	9.7	1079	5	09N4S7 caenorhabd
12	169.5	9.6	2045	11	061191 mus musculus
13	169.5	9.6	2045	11	09OJH2 mus musculus
14	167.5	9.4	549	5	061107 caenorhabd
15	166	9.3	514	4	09YX8 homo sapien
16	165.5	9.3	2187	11	P70670 mus musculus

17	164	9.2	1322	4	08TE50 homo sapien
18	164	9.2	1893	5	09NKC9 drosophila
19	160.5	9.1	886	12	09Q87 human herpe
20	159.5	9.0	2781	4	09UC2 homo sapien
21	159	9.0	513	4	043418 homo sapien
22	158.5	8.9	1844	5	022579 caenorhabd
23	157	8.9	847	5	019930 caenorhabd
24	157	8.9	3178	5	0969D4 caenorhabd
25	156.5	8.8	1696	4	08TE3 homo sapien
26	156	8.8	645	4	09H04 homo sapien
27	156	8.8	998	4	012869 homo sapien
28	156	8.8	1057	5	09W294 drosophila
29	156	8.8	1324	4	015312 homo sapien
30	155.5	8.8	842	5	09VGC9 drosophila
31	155.5	8.8	864	5	09VGC5 drosophila
32	155.5	8.8	864	5	09VGC8 drosophila
33	155.5	8.8	1151	13	057580 gallus gall
34	154.5	8.7	3507	5	023587 caenorhabd
35	153.5	8.7	2232	5	P91365 caenorhabd
36	152.5	8.6	577	13	0919D8 gallus gall
37	152.5	8.6	604	13	0919D7 gallus gall
38	151.5	8.5	961	3	092223 emericella
39	150	8.5	873	10	09C548 arabidopsis
40	149.5	8.4	1663	4	09UQ01 homo sapien
41	149.5	8.4	3261	4	09Y556 homo sapien
42	149.5	8.4	3664	4	096T58 homo sapien
43	149	8.4	586	13	091AB2 xenopus lae
44	147	8.3	744	3	08TRC9 schizosach
45	147	8.3	1029	4	08TDH7 homo sapien

ALIGNMENTS

RESULT 1

Q91W6 PRELIMINARY: PRT: 662 AA.
AC Q91W6; 01-DEC-2001 (TREMUREL, 19, Created)
DT 01-DEC-2001 (TREMUREL, 19, Last sequence update)
DT 01-JUN-2002 (TREMUREL, 21, Last annotation update)
DE TATA-binding protein associated factor TAFII135 (Fragment).
GN TAF4A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=PO BRAIN;
RA Metels M., Brunkhorst A., Neuman T.;
RT "Cell Type Specific Expression of the TFIID Component TAFII135 in the Nervous System."
RL Exp. Cell Res. 0:0-0(2001).
DR EMBL: AY038601; AAK94779.1; -
DR MGD; MGI:2152346; Taf4a.
FT NON_TER
SQ SEQUENCE 662 AA; 71398 MW; 91A75F38CBDD0DA4 CRC64;

Query Match 28.0%; Score 497; DB 11; Length 662;
Best Local Similarity 39.6%; Pred. No. 6.6e-26;
Matches 133; Conservative 53; Mismatches 102; Indels 48; Gaps 14;

QY 20 PRLAPQVAVKAKNTTIOFPANIQLPFGVILKSNGLMYSPOQTV-----TAAE 73
DB 10 PRLPQP-----QKPN-NIQ--NFQLPFGVILKSNGLMYSPOQTV-----TAAE 59
QY 74 TTSNITSPAVPAPQVYKIVTSPSSSOLIKKAVFPVKLIAIGTVVTVTPKPSVQ 133
DB 60 PQSITMAPRPAPTPGAPPPQISTVQAPGPITAR-QVTP-----TTIIOV---SQAQ 107
QY 134 SVAVPTSVVTVTPKPLNTVT--TLKPSGLASS-----TPSNE-PNKAENSAVQINTL 185

Db 108 TTVOPTTTLORSPGVQPOLVIGGSAQPASLGTAFTAVGTGTPORTVPGASTSTATE--- 164
Qy 186 SEPMLNKKCNFLAMLIKACSGSPBGMONKVLVEQLLDIAKEEFTKRLVEL 245
Db 165 ---TMBNKKCSFLSTLKLKSSGKOSTETANVNDLVONLLDGIKEADEFSRLREL 221
Qy 246 KSSPQHLVPEFLKSSVVALROLLPNSOSFIOCVOTSSDMVATCTTTPVTTTV 305
Db 222 NSSPQVLPFLKRSLLPALROLTPDSAAFIQSSQOQPP---ASQATTAITAVLSSSV 277
Qy 306 SSSQSEKSIIVSGATAPRTVSQVOTLPLAGPVGAKA 341
Db 278 QRTAGKTAASVTSALQPPVISTL--TQPTQVGVGKQA 311

RESULT 2

Q8T9E0 PRELIMINARY; PRT; 851 AA.
AC Q8T9E0;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE SD04735P.
GN TAF110.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OK NCBI_TaxId=7227;
RN
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokslein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guartin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Paclet J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,
RA Sumbit (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY069807; AAL39952.1;
SO SEQUENCE 851 AA; 92093 MW; 665B28B958C984C CRC64;

Query Match 11.0%; Score 194.5; DB 5; Length 851;
Best Local Similarity 26.9%; Pred. No. 3.4e-05;
Matches 101; Conservative 52; Mismatches 124; Indels 99; Gaps 19;

Qy 22 LPAPQIVANK--APNMTTQ--FPANQLPQGVLYIKNSGSLMLVSPQOTVRAETTSNI 78
Db 93 LPAGVYVGHROQAPSOQOQKNMPTN---PLSRVYINSHAGVAPQSP--SSMTTATTSNI 148
Qy 79 TSPAPVAPNPQTVKICTVPSNSS--OLIKKAVTPYKLAQIGTVVTTVPKPSVQSA 136
Db 149 I-----VNSVASSGVANSSOPHLVQLNMQAP--QLPQI--TQITIPAOQSOQ---- 193
Qy 13 VFTSVVTVTPGKPLNTVTTTLKPSISGASSTPSNEPLKAENSAVAOINLSPMLNENK K 195
Db 194 -----QOVNVVSSAGCTATVAVSSTTA-----ATTQOQWTKRK 226
Qy 196 CKNPLAMLIKILKLAESGSPBGMONKVLVEQLLDIAKEEFTKRLVELKSSPQHLVP 255
Db 227 CKFLANLIEL--STREKPEKENVKTLIOELVNAVNEEPCDRERLLNLSPOQCLIG 284
Qy 256 FLKSSVVALROL-----LPNSQSFIO-----QCVQO 281
Db 285 FLKSLPLRLRQALYTKELVIEGIKPPQHVLAGIAGLSQQLPKIQADIRIGISQTTTIG 344
Qy 282 TSSDMVATCTTTPVTTVSQVOTLPLAGPVGAKA 341
Db 345 TQVMI--TPNAGTPRPITGHTTISKOPPN--IRLPAPRLVNTGIRIQIP--SLQVP 398
Qy 339 AKAGVITLHSGPTAA 354
Db 399 GOANIVQIR--GPOHA 412

RESULT 3

Q76894 PRELIMINARY; PRT; 1795 AA.
AC Q76894;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE EG:5667.1 protein.
GN EG:5667.1 OR CG14796.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OK NCBI_TaxId=7227;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherzer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari P., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokslein P., Brotler P.,
RA Burks K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Donnes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durlin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,
RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kul P., Lai Z.,
RA Lasbo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlshina N.V., Moberly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN
RP SEQUENCE FROM N.A.
RA Cadieu E., Dreano S., Lelaure V., Mottier S., Gilbert F.;
RT "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RA Benos P.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE003421; AAF45644.1; -;
DR EMBL: AL031028; CAA19845.2; -;
DR FlyBase; FBgn0025390; EG:5667.1.
DR InterPro; IPR002357; Chitin_bind_Pera.
DR InterPro; IPR002965; P_rich_extensan.
DR Pfam; PF01607; CBM_14; 2.
DR PRINTS; PR01217; PRICHEXTENSAN.
DR SMART; SM00494; CHIBD2; 2.

DR MGD: MGI:10656; Zan.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR003645; FOLN.
 DR InterPro: IPR000988; MAM_domain.
 DR InterPro: IPR003328; TILA_Cysrich.
 DR InterPro: IPR002919; TILA_Cysrich.
 DR InterPro: IPR001007; VME-C.
 DR InterPro: IPR001846; VME-D.
 DR Pfam: PF00629; MAM; 3.
 DR Pfam: PF01826; TIL; 25.
 DR Pfam: PF02345; TILA; 25.
 DR Pfam: PF00094; vwd; 4.
 DR SMART: SM00001; EGF-like; 1.
 DR SMART: SM00274; FOLN; 21.
 DR SMART: SM00137; MAM; 3.
 DR SMART: SM00214; vwd; 25.
 DR SMART: SM00216; vwd; 4.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
 DR PROSITE: PS01186; EGF_2; 18.
 DR PROSITE: PS50060; MAM_2; 3.
 DR EGF-like domain; Glycoprotein.
 KW SEQUENCE 5374 AA; 579536 MW; 90D2D8CE5DE24EB CRC64;

Query Match 10.2%; Score 181; DB 11; Length 5374;
 Best Local Similarity 25.6%; Pred. No. 0.0025;
 Matches 95; Conservative 55; Mismatches 151; Indels 70; Gaps 19;

QY 2 TLVTKVAVSAPKPVSSGRLPAPQIVAVKAPMTTITQPPANLQLPQGVLIKNSGSLM 61
 DB 852 TEVAVTVLPASIPPEET-----TPTEVTTTPPEETI--PAEVTVP-----PVS 894
 QY 62 LVSPOQTVRAETTSNITSFANVPANQYKICTVPNSSQLIKKAAVTPVKLAQIGTT 121
 DB 895 IIS-EETTPTEVTVTPPEETIIPAEVTVVPVSIPEETTPTEVTVPEE-----TT 948
 QY 122 V--VTVPKPSSVOS--VAFTSVTVTPGKPL--NTVTLKPSLSGASTSPNEPNLK 174
 DB 949 IPAEVTVP-PVSIPEETIIPTEVTVPEETIIPAEVTVVPVSI-----PSEETI- 1001
 QY 175 AENSAVQINLSPTMLENKKCNFLAMLIKLAGSSQSPFEMQNVKKLEQLDAKIEA 234
 DB 1002 -----PTEVTVPEETIIPAEV-----TTPEETIIPTEVTVPPASISIP 1043
 QY 235 EETRLKLYELKSSPQHLVPLFKSVVALROLPPNSQSIQOC---VQOTS---SDMT 288
 DB 1044 EE-TASLFEVTVTPPEETIIPTEVTVPEETIIPAEV-----TTPEETIIPTEVTVPPASISIP 1102
 QY 289 ANCTTVTSPVYTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVAKAGVYTLHS 348
 DB 1103 ASEETIVSTQETTLTLEQSAVTQTSI-----ACRPQPS-PLMPTI-GPLSKPQVSMFS 1156
 QY 349 VGPFAVVGST 359
 DB 1157 LAPT--TGVT 1165

RESULT 6
 Q9JUK31 PRELIMINARY; PRT; 1306 AA.
 AC Q9JUK31;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE ATFA-associated factor.
 GN AT7IP OR AT7.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Degraeve F., Bahr A., Chation B., Kedinger C.;
 RT "A murine ATFA-associated factor with transcriptional repressing

RT activity."
 RL Ensembl: 10:1807-1819(2000).
 DR EMBL: AJ132702; CAB7024.1;
 DR MGD: MGI:1858965; At7f1p.
 DR InterPro: IPR002965; P-rich_extensn.
 DR PRINTS: PR01217; PRICEXTENS.
 SQ SEQUENCE 1306 AA; 138488 MW; 1P94ABA571D29191 CRC64;

Query Match 10.0%; Score 177.5; DB 11; Length 1306;
 Best Local Similarity 25.0%; Pred. No. 0.0081;
 Matches 108; Conservative 41; Mismatches 160; Indels 123; Gaps 19;

QY 5 TKVAVSAPKPVSSGRLPAPQIVAVKAPMTTITQPPANLQLPQGVLIKNSGSLM 61
 DB 761 TSTAIVSSQPKLQTSATSGSLPAPLPPASTATVAVTQVPSCTPQPTISLQPLVILH 820
 QY 33 -PNTTITQPPANLQLPQGVLIKNSGSLMIVSPQVTVRAETTSNITSFANVPANQYK 91
 DB 821 VPAVTVSQ-FQLQSHGTLVTNOPSQNVFISVQSPVTSGLTKNVSPLPL-NTKP 878
 QY 92 KICVPPNSQLIKKAAVTPVKLAQIGTV-VTVPKPSSVOSVAVPTSVTVTPG--- 147
 DB 879 NIPSVSPSS--IQNSSTA--APLGTLLAVQAVPTAHSIVQ-ARTSLPTVGPGLX 932
 QY 148 -----KPLN-----VTVTLKPSLSGASTSPNEPNLKENSANAQINLSPTMLENKKCN 198
 DB 933 SSSSSRGPIQMKIPISITFSPSS--SAEQNSSATPRIVAEQNTKTVDSIN-----KRAAD 986
 QY 199 FLAMLIKLAGSS-----QSPFEMQ--QNVKKLYQLDAKIEAEFTRKLYELKSS 248
 DB 987 STSOSGKASSDSGVDLIMDEESGTTQDPKISPPSSVTSQPMKRPLOPIIPAP 1046
 QY 249 P-OPHLVPLFKSVVALROLPPNSQSPFIQOCVQOTSQDMVIACTTVTSPVYTT- 304
 DB 1047 PLQPSGVPTSGPS-----QATIHVLPAPTVPVTRPVTQVYTRL 1087
 QY 305 -YSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVAKAGV----- 343
 DB 1088 PVPAFANHQVYVTLPLAPTTQA-----PLRGVMAQPAVAVQVNPONSVTVRPQTTYV 1142
 QY 344 ----VTLHSVGP 351
 DB 1143 VNNGLTIGSAGP 1154

RESULT 7
 Q9NVJ9 PRELIMINARY; PRT; 709 AA.
 AC Q9NVJ9;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CDNA FLJ10688 f1s, clone NT2RP300320, highly similar to partial
 DE putative p621 protein which interacts with transcription factor
 DE Sp1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
 RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
 RA Nakamura Y., Nagahara K., Masuhara Y., Sasaki N.;
 RT "NED0 human cDNA sequencing project."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK001350; BA91751.1;
 DR InterPro: IPR002965; P-rich_extensn.
 DR PRINTS: PR01217; PRICEXTENS.
 SQ SEQUENCE 709 AA; 75979 MW; F2C43814F6B241C9 CRC64;

Query Match 9.9%; Score 175.5; DB 4; Length 709;

Best Local Similarity 25.4%; Pred. No. 0.00053;
Matches 106; Conservative 53; Mismatches 166; Indels 93; Gaps 20;

QY 3 LTKVAPVAPK-----VSSGPRLPAPQIVAVKAPNTTTI-----QPAN-----43
164 LTPPRAVVSQPKLQTVTSG-SLTAISV--LAPNTATVATTQVSGNQPISLOPL 220
QY 44 -----LQDPGTVLKINSNCPMLVSPQOTVRAETTSNITSRAVAPAN 87
221 PVLHVPVAVSQPOLQSHPTLVNQPSGNVEFISVQSPVTSGLTKNPVSLPSLP-N 279
QY 88 POKVICTVNSSQILKKAIVPVKKLAOIGTV-VTVPKBSVOSVAVPTSVVTV 146
280 P--TKPNVSVSPSIOR--NPTASAPLGTTLAVQAVPTAHIAVO-ARTISLPTVGP 333
QY 147 GKPLNTVTTLP-----SSLGASTPSNEPKKENSAAVOINSPTLENVK 195
334 SGLSPSTNGPIQOMKIPISAFSTSSAABONSNTPTRIENQNTNTIASVSKKAADTSQ 393
QY 196 CKNFLAMLKLAGSGS-----PEMG--QNVKTLVEQLLDKIEAEFTRLKY 243
394 CGK-----ATGSDSGVIDLTMDDESGASQDKKLNHTPVSTMSSQPVSRLOP 444
QY 244 ELKSSP-QPHLVPFLKSVVALROLNPSQFIOQCVQOTSDDNVATCTTTVTSP--- 299
445 IQPAPPLQPSGVPPTSGPSQFTI-HLPTAPT---TVNVHPRVTVQVTRLVPRAANH 499
QY 300 -VTTTSSSSQSEKSI--IVSGATAPRTVSQVTLNPLAGPVGAKAGV---VTLHSVP 351
500 QVYTTTLPAPPAQPLAGTVMQAPAVQVNPQNSVTVRQTTTYVYVNNGLTLGTCG 557

RESULT 8
ID 090Y35 PRELIMINARY; PRT; 1979 AA.

AC 090Y35;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Host cell factor C1 (Fragment).
GN ABCD1 OR HCFC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Platter M., Brenner V., Reichwald K., Wiehe T., Oksche A.,
RA Rosenthal A.;
RT "Comparative sequence analysis of the mouse H1cam locus and the
RT corresponding region of human Xg28."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF133093; AAF22156.1;
DR MCD; MGI:105942; Hcfc1.
DR MCD; MGI:1349215; Abcd1.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF01344; Kelch; 4.
DR SMART: SM00060; FN3; 1.
FT NON_TER 1
SQ SEQUENCE 1979 AA; 203436 MW; ABE998C5DDE688A9 CRC64;

Query Match 9.8%; Score 174.5; DB 11; Length 1979;
Best Local Similarity 23.6%; Pred. No. 0.0021;
Matches 106; Conservative 60; Mismatches 144; Indels 139; Gaps 22;

QY 10 VSAAPPVSSPRLPAPQIVAVKAPNTTTIOPANLQPLPGTVLKNSGMLVSPQOTV 69
471 IGSNPKSGMAALAAAAAATOKIPSSA---PVLVSPAGTIVKT-----VAVTPGTT 522
QY 70 TRAFETSNITSRAVAPANPOT-----VAKCTVPSSSQ---LKKVAVTPVKKL 115

Db 523 LPA--TVKVASPVWVSNPATRMLKTAAGVTSVSSAANTSTRPITTVHKSQTVTAQ 580
QY 116 AQIGTVV-----TVPKP-----SSQVAVPTS-----140
Db 581 AQVTTVVGATKTIITLVKSPISVPGSALISNGKWSVQTPVQTSNAGASTGPV 640
QY 141 -----VTVTPGKPLNTVTT-----LKPSSLAGAST-----PSNE 170
Db 641 TQIOTKGPLPAGTILKLVTSADGKPTTIITTTQASGAGTPTLIGSSVSPSTTKPCT 700
QY 171 PNLAKENSAVOINLSPMLENVKCKNFLML-IKLACSSQSPKMGQNVKLVLEOLD 229
701 TIKTIPMSALITQAGAGVTSVPGIKSPITITTKVTSQTAGP-----AKITAV-- 752
QY 230 AKIEAEFTRLK-VYELKSSP-QP-----HLVPFLKSVVALROLNPSQFIOQCVQOTS 283
753 PKITGHCQCVTVQVLKAGQPGTILIRYP-----MGRVLVTPVTSVAVKAV---- 803
QY 284 SDNVIANCTVTTVTSPPVTTVSSQSEKSIIVSGATAPRTVSQVTLNPLAGPVGAKGV 343
804 -TTLVKGTTGVTTLGTVTGVTS-----LAGAGHSTSA-----SLATPITLGTI 850
QY 344 VTLHS--VGPTA-----ATGCTT 359
Db 851 ATLSQVINPPIATVSAAGTTLTAAGLT 879

RESULT 9
ID 029071 PRELIMINARY; PRT; 528 AA.

AC 029071;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Gastric mucin (Fragment).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE-GASTRIC MUCOSA;
RX MEDLINE=95275264; Pubmed=755593;
RA Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
RA Lamont J.T.;
RT "Isolation and characterization of cDNA clones encoding pig gastric
RT mucin."
RL Biochem. J. 308:89-96(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE-GASTRIC MUCOSA;
RX MEDLINE=94102478; Pubmed=7506218;
RA Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
RA Lamont J.T.;
RT "Pig gastric mucin: isolation and characterization of a cDNA clone
RT with a novel tandem repeat."
RL Gastroenterology 106:200-200(1994).
DR EMBL: U10281; AAC48526.1;
FT NON_TER 1
FT NON_TER 528
SQ SEQUENCE 528 AA; 49907 MW; 0BF0F6879203B2EA CRC64;

Query Match 9.8%; Score 173.5; DB 6; Length 528;
Best Local Similarity 23.3%; Pred. No. 0.00051;
Matches 97; Conservative 61; Mismatches 138; Indels 121; Gaps 20;

QY 8 APVSAAPPV--SSGPRLPAPQIVAVK-----APNT--TTIOPANLQPLP--GTVLK 55
108 APISSTSVQSSGSGVPTTSATSVQSSSSSAPVTSATSVQ--PSSSSSPISSTVSVQ 166
QY 56 NSGPMLVSPQOTVTRAFETSNITSRAVAPANPOTVAKICVPSNSQILKKAIVTPVK-- 113
Db 167 SSS-----SSAPVTSATSVQSPSSSSSPISSTVSVQTSSTTSVQPS 215

```

0Y 114 KLAOGITVVVVY-----PKPSSVO-----SVAPVSVVYVEGK----PLNVVT 155
Db 216 SSSSVPTTSATSVRRSSSSSTPIPTSTSVQSSSSSAPPTTSATSVQSSSSSTPIPTTS 275
0Y 156 LKPSLIGASSTP---SNEPNLKAENSAVOINLSEPTLMEVKKCNFLMLIKIACSG 210
Db 276 VOPSS--SSSAPPTSATSVOPSSSSSPISSTIVOS-----SS 313
0Y 211 SQSPMGONVKKIVQLLDAKIEAEFFRKILYELKSSPOHVLVPLFKSVVALROLPLN 270
Db 314 SSSPTTS-----TTSVOPSSSGSAPPTTSATSVOPSSSSSP1-----SSTISVO--PS 359
0Y 271 SOS-----FIO-----OCVOQTSADMVIACTTIVTS-----PVYTVV 305
Db 360 SSSSPTTSTSVQPOSSSGSAPPTSVANSVQSSSSSVPTTSATSVRRSSSSSTPIPTTS 419
0Y 306 SSSOSKEIIVSGANA-----PRYSVOTLNLPLAGPVCAKKGVTLSHSGP 351
Db 420 VOPSSSSVPTTSATSVQSSSSSTPIPTSTSVQSSSSSAPPTTSATSVOPSSSSSP 476

```

ID	OBMMQ4:	PRELIMINARY:	PRT:	1349 AA.
AC	OBMMQ4:			
DT	01-MAR-2002 (TREMBLrel. 20, Created)			
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	Mucin 5 (Fragment).			
GN	MUC5AC.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_Taxid=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-21426417; PubMed-11535137;			
RA	Escande F., Aubert J.P., Porchet N., Buisine M.P.;			
RT	"Human mucin gene MUC5AC: organization of its 5'-region and central			
RT	repetitive region.";			
RL	Biochem. J. 358:763-772(2001).			
DR	EMBL: A1298318; CAC83675.1; -.			
FT	NON_TER	1		
FT	NON_TER	1349	1349	
SO	SEQUENCE	1349 AA;	135600 MW;	4DC3C1544F1E5EBA CRC64;
Query Match 9.7%: Score 172.5; DB 4; Length 1349;				
Best Local Similarity 22.3%: Pred. NO. 0.0018;				
Matches 94; Conservative 55; Mismatches 178; Indels 95; Gaps 13.				
QY	1	GLVTVKAVPVSAPPKVVSGGRLLP-----APQIVAAKAPNTWTI	38	
DB	301	GCPTVS-PTVTAAPSPSGAATSPQTSSQMSKRTTLTVTSTTSPQSTISAPTSTT	359	
QY	39	QFPANIQLPFGVGLIKNSGPLMLVSPQOTVRAETTSNITSRPAVPANPQVTKICTYVN	98	
DB	360	SAPPTSTSPAPTSTTSTPQTSISSAPTSSTTSPAPTSTISAR-----TTSITAPT	411	
QY	99	SASQILKKAIVPVKKLQIGTTVTVTKKPSVQS-----VAVPISVTVTPGKPLNTV	153	
DB	412	TST-----TSSPTTSTTSATTTSTTSAPTSTTSPQTSKTSAAITSTSSGTTTPSPV	465	
QY	154	TLTKPSSLGASSTPSNEPLKAENSAAYQINLSP-----	187	
DB	466	TTTSTASVSKIT- SHVSVSKTHSHQPYTRDCHPCTYTKWFDVDFPSPGPGDKETYN	524	
QY	188	TLNLEVKKCKNLFAMLIKLAGSGSOSPE-----MGQNVKYLEQLLDAKIEAEFFTKRL-	241	
DB	525	NIRSGEIKCRPQETTRIQCRAKSHPEVSIHLGQVCGCSREGLVCNQQQGPFKMC	584	
QY	242	---YVELSSPQPHLVLPFLKKSVALIRQLL--PNSQSTQCCVQOTSSDMVATCTTVT	296	

```

Db      585  INYEVRVLCGETPKGCCPYTSTSVAPSLVINGEPPAQI-----QSTSMQ- -KSRITTLV 636
QY      297  TSPVVTTSVSSQSEKSLIVGCAAPRVSVQTLNPLAGPYGAKAGAVTILHSVGNPAATG 356
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      637  TTSIISTQIQTSTTAPITSTTPASIPSTTSAPTTSTTSAP-----TSTTSAPTTST 689
QY      357  GT 358
        :
Db      690  ST 691

```

RESULT 11		
Q9N4S7		
ID	Q9N4S7	PRELIMINARY; PRT; 1079 AA.
AC	Q9N4S7;	
DT	01-OCT-2000 (TrEMBLrel. 15, Created)	
DR	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)	
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)	
DE	Y51B1A.1 protein.	
GN	Y51B1A.1.	
OS	Caenorhabditis elegans.	
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae	
OC	Rhabditidae; Peloderiinae; Caenorhabditis.	
OX	NCBI_TaxID=6239.	

RC	STRAIN-BRISTOL N2;
RX	MEDLINE-99069613; PubMed-9851916;
RA	None;
RT	"Genome sequence of the nematode C. elegans: a platform for
RT	investigating biology. The C. elegans Sequencing Consortium.";
RL	Science 282:2012-2018(1998).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN-BRISTOL N2;
RA	Cotton M.;
RT	"The sequence of C. elegans cosmid Y51B11A.";
RL	Submitted (MAR-1999) to the EMBL/Genbank/DDbj databases.
RN	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN-BRISTOL N2;
RA	Waterston R.;
RL	Submitted (MAR-2000) to the EMBL/Genbank/DDbj databases.
DR	EMBL, AC006797; AAF60743.1; -
DR	InterPro; IPR002965; P-rich_extensn.
DR	PRINTS; PR01217; PRICHEXTENS.
SQ	SEQUENCE 1079 AA; 110532 MW; 8DBDE3824CF80CA1 CRC64;
 Query Match 9.7%; Score 171.5; DB 5; Length 1079;	
Best Local Similarity 22.6%; Pred. No. 0.0016;	
Matches 82; Conservative 38; Mismatches 169; Indels 74; Gaps 10.	
OY	5 TKNAPVSA---PPKVSSEGLRPAPQIVAKABNTTITOPANIQLPPTVL-----IKSN 56
Dd	399 TTAAETTTSTEPSSTTP-----VQTITTAAETTSTEPSSTTPVQTTTTTAETTST 454
OY	57 SGPMLAVSQGTATRA--ETTSNITSRPAPVPANPQTVKICTVPNSSQLIKKVAATPYKK 114
Dd	455 EPPSSSTTPVQTTTTTAETTST--TEPSSSTTPVQTTTTTAETTSTEPSSTTPVQ- 511
OY	115 LAQIGTIVVTYPKRSSVQSVAVPFISVTVTGCKPLNTYTTLKPSS-----LGASS 165
Dd	512 -----TTTTTAETTSTEPSSTSTTPVQTTTTTAETTSTEPSSTTPVQTTTTTAPE 565
OY	166 TPSPENPKAENSAVNOILSPTMLENVKKCNFLAMLIKLAGSGSQSPDMQNKKIYE 225
Dd	566 TTSTEP----PSSSTTPVQTTTTTAETTSTEPSSTSTTPVQTTTTTAETTST----- 615
OY	226 QLDARKIEAEETTRKLIVELKSSPOPHLVPEFKKSVALROLPLPSQSEIQQCVQOTSDD 285
Dd	616 -----EPSSSTTPVQTTTTTAETTSTEPSSTSTTPVQTTTTT 654
OY	286 MYATACT-----TTVTSFPVTTTVSSSQSESKIIVSGATNPATPVSVQTLNPLA 334

```

Db 655 APEETSTPPSSSTPVQTTTAAETTTSTPPSSSTPVQTTTAAETTTSTPPSSST 714
QY 335 GPV 337
Db 715 SPV 717

RESULT 12
Q61191
ID 061191 PRELIMINARY; PRT; 2045 AA.
AC 061191:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Transcription factor C1 (HCF).
GN HCF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Kristie T.M.; Dasher R.;
RT "CDNAs encoding the mouse homolog of the human transcription factor C1
RT (HCF).";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U53925; AAB01163.1; -.
DR MGD: MGI:105942; Hcfcl.
DR InterPro: IPR003961; FN.III.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF01344; Kelch; 5.
DR SMART: SM00060; FN3; 1.
SQ SEQUENCE 2045 AA; 210537 MW; B89CC2FDA35B969F CRC64;

Query Match 9.68; Score 169.5; DB 11; Length 2045;
Best Local Similarity 23.48; Pred. No. 0.0048;
Matches 105; Conservative 60; Mismatches 145; Indels 139; Gaps 22;

QY 10 VSAPPKVSRRPAPQIVAVKAPNTTIOFPANQLPFGVILKSNGLMLVSPQOTV 69
Db 537 IGSNPQSGMALALAAATAOKIPSSA---PTAMSVAGTTIYKT-----VAATPGT 588
QY 70 TRAEITNITSRPVAVPAPNPT-----VKICTVPNSSO---LIKKVAATPVKRL 115
Db 589 LPA--TVKAVSSPVVSNPATRMLKTAAGVGTSSAANTSTRPIITVHKSGVTVAQ 646
QY 116 AQIGTIVV-----TVPKP-----SSVQSAVPTS----- 140
Db 647 AQVTTVVGVTKITLVKSPISVPGSALISNLKVMVSVQTRKVOGSAVTAQASTGPV 706
QY 141 -----VVTTPGKPLNTVTT-----LKPSIGAST-----PSNE 170
Db 707 TQIIHTKPLPAGTILKLVTSADGKPTTIIITTTQASGAGTPTIIGISVSPSTTKPGTT 766
QY 171 PNLKAENSAVQINLSPTMLENVKCKNFLLML-IKLACSGSQSPENQNVKRLVEQLLD 229
Db 767 TIITIKPSALITQAGATVTSPPGKSPITIIITTKVMTSGTGA-----AKITAV-- 818
QY 230 AKIABEFTKRL-YVELKSSP-QP---HLVPFLAKSVVALROLPLNSQSFIOQVOOTS 283
Db 819 PKIATGHGQCGVTOVLKAGAPGPGTILRTVP-----MGVRLVTPVVSAAVKPRAV---- 869
QY 284 SDMYATCTTIVTSPVTTVSSQSEKSIIVSGATAPRTVSVQTLNPLAGPVGAKAGV 343
Db 870 -TTLVVKGTTGVTGLVGTGIVTS-----LAGAGHSTSA-----SLATPIYTLGCTI 916
QY 344 VTLHS--VGPTA-----ATGTT 359
Db 917 ATLSQVINPFAITVSAAGTTLTAAGLIT 945

```

```

RESULT 13
Q61707
ID 061707 PRELIMINARY; PRT; 2045 AA.
AC 061707:
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE C1 transcription factor.
GN HCF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC Kristie T.M.;
RT "CDNAs encoding the mouse homolog of the human transcription factor C1
RT (HCF).";
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U80821; AAD09225.1; -.
DR MGD: MGI:105942; Hcfcl.
DR InterPro: IPR003961; FN.III.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF01344; Kelch; 5.
DR SMART: SM00060; FN3; 1.
SQ SEQUENCE 2045 AA; 210546 MW; 7AD38FCD78EABA9C CRC64;

Query Match 9.68; Score 169.5; DB 11; Length 2045;
Best Local Similarity 23.48; Pred. No. 0.0048;
Matches 105; Conservative 60; Mismatches 145; Indels 139; Gaps 22;

QY 10 VSAPPKVSRRPAPQIVAVKAPNTTIOFPANQLPFGVILKSNGLMLVSPQOTV 69
Db 537 IGSNPQSGMALALAAATAOKIPSSA---PTAMSVAGTTIYKT-----VAATPGT 588
QY 70 TRAEITNITSRPVAVPAPNPT-----VKICTVPNSSO---LIKKVAATPVKRL 115
Db 589 LPA--TVKAVSSPVVSNPATRMLKTAAGVGTSSAANTSTRPIITVHKSGVTVAQ 646
QY 116 AQIGTIVV-----TVPKP-----SSVQSAVPTS----- 140
Db 647 AQVTTVVGVTKITLVKSPISVPGSALISNLKVMVSVQTRKVOGSAVTAQASTGPV 706
QY 141 -----VVTTPGKPLNTVTT-----LKPSIGAST-----PSNE 170
Db 707 TQIIHTKPLPAGTILKLVTSADGKPTTIIITTTQASGAGTPTIIGISVSPSTTKPGTT 766
QY 171 PNLKAENSAVQINLSPTMLENVKCKNFLLML-IKLACSGSQSPENQNVKRLVEQLLD 229
Db 767 TIITIKPSALITQAGATVTSPPGKSPITIIITTKVMTSGTGA-----AKITAV-- 818
QY 230 AKIABEFTKRL-YVELKSSP-QP---HLVPFLAKSVVALROLPLNSQSFIOQVOOTS 283
Db 819 PKIATGHGQCGVTOVLKAGAPGPGTILRTVP-----MGVRLVTPVVSAAVKPRAV---- 869
QY 284 SDMYATCTTIVTSPVTTVSSQSEKSIIVSGATAPRTVSVQTLNPLAGPVGAKAGV 343
Db 870 -TTLVVKGTTGVTGLVGTGIVTS-----LAGAGHSTSA-----SLATPIYTLGCTI 916
QY 344 VTLHS--VGPTA-----ATGTT 359
Db 917 ATLSQVINPFAITVSAAGTTLTAAGLIT 945

RESULT 14
Q61707
ID 061707 PRELIMINARY; PRT; 549 AA.
AC 061707:
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, last annotation update)

```

DE Hypothetical 60.4 kDa protein.
GN R119.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=9069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RM [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Clarke K., Bauer C., O'Neal D.;
RT "The sequence of C. elegans cosmid R119.";
RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
RM [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct submission.";
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF063007; AAC16427.1; -;
DR InterPro; IPR004822; Histone_core.
DR InterPro; IPR003894; TAF_hom.
DR SMART; SM00549; TAFH; 1.
KW Hypothetical protein.
SQ SEQUENCE 549 AA; 60430 MW; 17AF21E0937B9B38 CRC64;
Query Match 9.4%; Score 167.5; DB 5; Length 549;
Best Local Similarity 28.1%; Pred. No. 0.0014;
Matches 63; Conservative 27; Mismatches 91; Indels 43; Gaps 8;
QY 158 PSLGASSTPSPNPNKAENSAVQINLSPTMEENYKCKNFMAMLIKACSSQSPEN- 216
DB 102 PSPMAPQPGPT-----PQNSAAAAAASDD--KNYTKCVRFKTLTINT--SNNDPEMP 152
QY 217 --GQNYKKLVEQLDAKIAEAEFTRLKLYELKSSPOPHLVPLFKSSVALROLPLNSQSF 274
DB 153 DKAARKYELIRGIVYLETAEETRLQOVLKSAQPHLPLQNTLPLALNRNRGTAS 212
QY 275 I-----GQCVOQTSSDMVIANCTVTTSPTVTTSSS 308
DB 213 VEGVNPPEGVTFNNGRTPPGPQPPPOQSQOQPLEMQIIPPNQIPQMVGGPHMVS 272
QY 309 QSEKSTI-VSGATAPRTVSQTLNPLAGPVGAKAGVTLHSVCP 351
DB 273 VGARPMIRPMGPGGSPMGIQ--GPGRGPMGHO--VMQMHPPP 312
RESULT 15
QY4X8
ID QY4X8 PRELIMINARY; PRT; 514 AA.
AC QY4X8;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE P621 (Fragment).
GN 621.
OS Homo sapiens (Human).
OG Plasmid pectil.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-COLON;
RX MEDLINE=20431278; PubMed=10976766;
RA Gunther M., Lathier M., Brison O.;

RT "A set of proteins interacting with transcription factor Spl
identified in a two-hybrid screening.";
RL Mol. Cell. Biochem. 210:131-142(2000).
DR EMBL; AJ242978; CAB45135.1; -;
DR InterPro; IPR002965; P-rich_extensn.
DR PRINTS; PR01217; PRICHEXTENSX.
KW Plasmid.
FT NON_TER 1 1
FT NON_TER 514 514
SQ SEQUENCE 514 AA; 54963 MW; D4C5E5AC598E1DBA CRC64;
Query Match 9.4%; Score 166; DB 4; Length 514;
Best Local Similarity 25.1%; Pred. No. 0.0016;
Matches 91; Conservative 41; Mismatches 132; Indels 98; Gaps 17;
QY 3 LVTKVAVSAPK-----VSSGPRLPAPQIVAVKAPNTTTI-----QPPAN----- 43
DB 181 LVTPPNAVSSQPKLQPTVTG-SLTAISV--LRPNATVYATTOVPSGNPOTISLOPL 237
QY 44 -----LQPPGVTLKNSGPLMLVSPQOTVRAETTSNITSRPAPVAN 87
DB 238 PVILHVPVAVSSQPOLQSHPGTLVTNQPGNVEFISVQSPPTVSGLTKNPVSLPLP-N 296
QY 88 POTVKICTVPSNSSQLKKYAVTPVKLAIGTIV-VTVPKFSSVQSAVPTSVTVTP 146
DB 297 P--TKPNVPSVSPSTIQ--NPTASAPLGGTLAVQAVPTAHSIVQ-ATRTSLPTVGP 350
QY 147 GKPLNTVTTLKP-----SSLGASSTSPNEPMLKAENSAVQINSPMLENVK 195
DB 351 SGLYSPSTNGPQOMKIPISAFSTSSAAEQNSMTPTRIENQTKTIDASVSKRAADTSQ 410
QY 196 CKNFAMLIKACSSQS-----PEMG--QNYKKLVEQLDAKIAEFTRLYV 243
DB 411 CGK-----ATGSDSGVIDLTMDDESGASQDPRKLNHTIPVSTWSSSQPVSRPLQ 461
QY 244 ELKSSP-QPHLVPEFKKSSVALROLPLNSQSFIOQCVOQTSMDVIANCTVTTSPTV 302
DB 462 IQPAPPLQPSGVPTSGPSQTTI-HLPT-----APTVNVTHRVYQ 502
QY 303 TT 304
DB 503 VT 504
Search completed: February 16, 2003, 21:59:10
Job time : 33.2146 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:40:36 ; Search time 34.6501 seconds
(without alignments)
1738.213 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_452

Perfect score: 2249
Sequence: 1 GLVTKVAPVSAAPRVSSGP.....VIGTPVQIKIAPGPVLSQP 452

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: A.Geneseq_101002.*
2: /SID2/gcgdata/geneseq/genesqp-emb1/AA1980.DAT:*
3: /SID2/gcgdata/geneseq/genesqp-emb1/AA1981.DAT:*
4: /SID2/gcgdata/geneseq/genesqp-emb1/AA1982.DAT:*
5: /SID2/gcgdata/geneseq/genesqp-emb1/AA1983.DAT:*
6: /SID2/gcgdata/geneseq/genesqp-emb1/AA1984.DAT:*
7: /SID2/gcgdata/geneseq/genesqp-emb1/AA1985.DAT:*
8: /SID2/gcgdata/geneseq/genesqp-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/genesqp-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/genesqp-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/genesqp-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/genesqp-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/genesqp-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/genesqp-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/genesqp-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/genesqp-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/genesqp-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/genesqp-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/genesqp-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/genesqp-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/genesqp-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/genesqp-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/genesqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2249	100.0	801	19	AAW31494
2	2249	100.0	801	22	ABG09468
3	2249	100.0	852	21	AAI57279
4	521	23.2	737	15	AAW56494
5	521	23.2	737	17	AAW06084
6	521	23.2	737	13	AAW25019
7	513	22.8	1023	23	AAW82954
8	241.5	10.7	921	15	AAW56487
9	241.5	10.7	921	17	AAW06077
10	241.5	10.7	921	18	AAW25028

11	241.5	10.7	921	22	ABB61528
12	241.5	10.7	921	12	ABW66055
13	204	9.1	2035	15	AAW57141
14	202	9.0	1795	22	ABW69806
15	199.5	8.9	842	22	ABW66631
16	199.5	8.9	864	22	ABW71319
17	195.5	8.7	5179	22	AAW24516
18	191	8.5	1930	22	AAW65656
19	191	8.5	2135	23	AAW21714
20	181.5	8.1	571	22	AAW27242
21	181.5	8.1	708	22	AAW79978
22	181.5	8.1	709	22	AAW78994
23	181.5	8.1	709	22	AAW93063
24	181.5	8.1	709	22	AAW94334
25	181.5	8.1	881	22	AAW93811
26	181	8.0	1296	23	ABW66702
27	181	8.0	1296	23	ABW66756
28	180.5	8.0	1236	23	AAW42650
29	178	7.9	1127	22	AAW95541
30	177.5	7.9	662	22	ABW71837
31	176.5	7.8	1714	22	ABW60186
32	176.5	7.8	2972	22	AAW50363
33	176.5	7.8	3118	22	AAW50362
34	175	7.8	1658	22	ABW67620
35	173.5	7.7	557	22	AAW94078
36	173.5	7.7	849	17	AAW06725
37	173.5	7.7	2781	21	AAW57453
38	173.5	7.7	2870	21	AAW95559
39	173.5	7.7	2907	21	AAW57452
40	173.5	7.7	3178	21	AAW95556
41	172	7.6	750	20	AAW05477
42	171.5	7.6	1721	19	AAW48299
43	171.5	7.6	1721	21	AAW11727
44	171.5	7.6	1721	21	ABJ04045
45	169.5	7.5	1057	23	ABW65440

ALIGNMENTS

RESULT 1	
AAW31494	
ID	AAW31494 standard; Protein: 801 AA.
XX	
AC	AAW31494;
XX	
DT	28-APR-1998 (first entry)
XX	
DE	Human hTAFII105 protein.
XX	
KW	TAFII-binding protein associated binding factor 105; human; activator;
KW	hTAFII105; transcription factor; TFIID; transcriptional activation;
KW	antibodies; diagnosis; therapy; biopharmaceutical industry.
XX	
OS	Homo sapiens.
XX	
PN	US5710025-A.
PD	20-JAN-1998.
XX	
PF	02-OCT-1996; 96US-0725012.
XX	
PR	02-OCT-1996; 96US-0725012.
XX	
PA	(REBC) UNIV CALIFORNIA.
XX	
PI	Dikstein R, Tjian R;
XX	
DR	WPI: 1998-109818/10.
XX	
PT	N-PSDB; AAV02872.
XX	
PT	DNA encoding human tata-binding protein associated factor - for
XX	producing recombinant protein

XX Claim 1; Col 17-22; 12pp; English.
PS
XX This cDNA sequence represents a human tata-binding protein associated
CC factor, htraf1105, isolated from Daudi cell nuclear extracts. Tightly
CC associated subunits (traf's) are components of the transcription factor
CC TRAF1 and are thought to mediate transcriptional activation. This encoded
CC protein may be produced recombinantly from transformed host cells or
CC purified from human cells. htraf1105 specific binding agents such as
CC specific antibodies could be used for diagnosis (e.g. genetic
CC hybridisation screens for htraf1105 transcripts), therapy (e.g. gene
CC therapy to modulate htraf1105 gene expression) and in the
CC biopharmaceutical industry (e.g. as immunogens, reagents for isolating B
CC cell specific activators or other transcriptional regulators).
XX
SQ Sequence 801 AA:
Query Match 100.0%; Score 2249; DB 19; Length 801;
Best Local Similarity 100.0%; Pred. No. 8,3e-159;
Matches 452; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTLVTKVAVSAPPKSSGPRLPAPQIVAVKAPNTTIOFPANLQLPPTGLIKNSGFL 60
DB 1 GTLVTKVAVSAPPKSSGPRLPAPQIVAVKAPNTTIOFPANLQLPPTGLIKNSGFL 60
QY 61 MLVSPQQTIVTRAETTSNITSRPAVPANPQTVKICVTPNSSSOLIKKVAATPVKKLAQIGT 120
DB 61 MLVSPQQTIVTRAETTSNITSRPAVPANPQTVKICVTPNSSSOLIKKVAATPVKKLAQIGT 120
QY 121 TVTWTVPKSSVQSVAVPSTVTPGKPLNTVTLKPSLSGASSTPSENPMLKAKNSAA 180
DB 121 TVTWTVPKSSVQSVAVPSTVTPGKPLNTVTLKPSLSGASSTPSENPMLKAKNSAA 180
QY 181 VOINISPTMLENVKCKKNFLAMLIKACSGSOSPEMGQVKKLVQDLDAKTEAEFEFTKR 240
DB 181 VOINISPTMLENVKCKKNFLAMLIKACSGSOSPEMGQVKKLVQDLDAKTEAEFEFTKR 240
QY 241 LVELKSSPQPHLPFLKRSVVALRQLPNSQSFIOQCVOQTSDBVATCTTWTTSVP 300
DB 241 LVELKSSPQPHLPFLKRSVVALRQLPNSQSFIOQCVOQTSDBVATCTTWTTSVP 300
QY 301 VTTVSSSSOSEKSIIVSGATAPRTVSVQTLNPLAGPVGAKAGVTLHSGPTAATGCTTA 360
DB 301 VTTVSSSSOSEKSIIVSGATAPRTVSVQTLNPLAGPVGAKAGVTLHSGPTAATGCTTA 360
QY 361 GTGLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAVTSLPVTGEGTSGAAILCLPSVKP 420
DB 361 GTGLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAVTSLPVTGEGTSGAAILCLPSVKP 420
QY 421 VVSFCMDHICKPVIGTPVOIKLAOPGVLSOP 452
DB 421 VVSFCMDHICKPVIGTPVOIKLAOPGVLSOP 452
RESULT 2
ABG09468 ID ABG09468 standard; Protein: 801 AA.
AC ABC09468;
DT 13-FEB-2002 (first entry)
DE Novel human diagnostic protein #9459.
XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS73655.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
PS Claim 20; SEQ ID NO 39827; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC (polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 801 AA:
Query Match 100.0%; Score 2249; DB 22; Length 801;
Best Local Similarity 100.0%; Pred. No. 8,3e-159;
Matches 452; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTLVTKVAVSAPPKSSGPRLPAPQIVAVKAPNTTIOFPANLQLPPTGLIKNSGFL 60
DB 1 GTLVTKVAVSAPPKSSGPRLPAPQIVAVKAPNTTIOFPANLQLPPTGLIKNSGFL 60
QY 61 MLVSPQQTIVTRAETTSNITSRPAVPANPQTVKICVTPNSSSOLIKKVAATPVKKLAQIGT 120
DB 61 MLVSPQQTIVTRAETTSNITSRPAVPANPQTVKICVTPNSSSOLIKKVAATPVKKLAQIGT 120
QY 121 TVTWTVPKSSVQSVAVPSTVTPGKPLNTVTLKPSLSGASSTPSENPMLKAKNSAA 180
DB 121 TVTWTVPKSSVQSVAVPSTVTPGKPLNTVTLKPSLSGASSTPSENPMLKAKNSAA 180
QY 181 VOINISPTMLENVKCKKNFLAMLIKACSGSOSPEMGQVKKLVQDLDAKTEAEFEFTKR 240
DB 181 VOINISPTMLENVKCKKNFLAMLIKACSGSOSPEMGQVKKLVQDLDAKTEAEFEFTKR 240
QY 241 LVELKSSPQPHLPFLKRSVVALRQLPNSQSFIOQCVOQTSDBVATCTTWTTSVP 300
DB 241 LVELKSSPQPHLPFLKRSVVALRQLPNSQSFIOQCVOQTSDBVATCTTWTTSVP 300
QY 301 VTTVSSSSOSEKSIIVSGATAPRTVSVQTLNPLAGPVGAKAGVTLHSGPTAATGCTTA 360
DB 301 VTTVSSSSOSEKSIIVSGATAPRTVSVQTLNPLAGPVGAKAGVTLHSGPTAATGCTTA 360
QY 361 GTGLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAVTSLPVTGEGTSGAAILCLPSVKP 420
DB 361 GTGLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAVTSLPVTGEGTSGAAILCLPSVKP 420

Db	361	GTGILGTQSKPLVTSVANTVTWVSLQPERKRVSGTAVTSLSPAVTGETSGAICLPVSK	420
Qy	421	VVSEFCMDHICKPVITGTPVOIKIAGGEPVLSQ	452
Db	421	VVSEFCMDHICKPVIGTPVOIKIAGGEPVLSQ	452
RESULT 3			
ID	AAV57279	standard; Protein, 852 AA.	
XX	AAV57279;		
XX	06-JUN-2000	(first entry)	
DE		Transcription factor subunit TAFII105 polypeptide.	
XX			
KW		TATA box-binding protein associated factor II 105; TAFII105; cancer;	
KW		transcription factor; apoptosis; cytostatic; immunosuppressive;	
KW		antiinflammatory; vitruide; antibacterial.	
XX			
OS		Homo sapiens.	
XX			
FH	Key	Location/Qualifiers	
FT	Misc-difference	802	
FT		/label= Glx	
FT		/note= "encoded by TGA"	
FT	Misc-difference	834	
FT		/label= Glx	
FT		/note= "encoded by TAA"	
XX			
PN	WO200012699-A1.		
XX			
PD	09-MAR-2000.		
XX			
PE	25-AUG-1999;	99WO-IL00464.	
XX			
PR	27-AUG-1998;	98IL-012597L.	
XX			
PA	(YEDA) YEDA RES & DEV CO LTD.		
XX			
PI	Dikstein R, Yamit-hezi A;		
DR	WPI: 2000-256640/22.		
DR	N-PSDB; AA290465.		
XX			
PT	Polypeptide encoding TATA box binding protein associated factor II 105		
PT	useful for treating e.g. cancers and inducing apoptosis has a dominant		
PT	negative effect on the normal biological activity of the binding		
PT	protein -		
XX			
PS	Claim 7; Fig 2; 48pp; English.		
XX			
CC	This represents a polypeptide comprising a (modified) fragment (I) of		
CC	a TATA box-binding protein associated factor II 105 (TAFII105). A		
CC	pharmaceutical composition comprising (I) or the polynucleotide or an		
CC	inhibitor or antagonist of (I) is useful for treating cancers and		
CC	inducing apoptosis in pathological cells. The composition is also useful		
CC	for treating autoimmune diseases, inflammatory processes and viral or		
CC	bacterial infections.		
XX			
XX	Sequence	852 AA;	
Qy	Query Match	100.0%; Score 2249; DB 21; Length 852;	
Db	Best Local Similarity	100.0%; Pred. No. 9e-159;	
	Matches	452; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
Qy	1	GLTVTKVAPVSAAPRVSSGPRLLPAPQIVAVKAPNTTITQFPANLQLPQGTVLIKSNSGPL	60
Db	1	GLTVTKVAPVSAAPRVSSGPRLLPAPQIVAVKAPNTTITQFPANLQLPQGTVLIKSNSGPL	60
Qy	61	MLVSPQGVYTRAFETSNITSRPVAVPANTQVATCTVPPNSSQILIKKAVATPVKKLAQIST	120

Db	61	MLVSPQGVYTAETTSNTSIRAPVAPANQYKICIVPSSSOLIKKVAVTYVKLLAQIGT	120
Oy	121	TVVTVTPKRPSSVQSAVAPTVSVTVTPGKPLMTVTTLKRPSSLGCASTPSENPDLKAENSA	180
Db	121	TVVTVTPKRPSSVQSAVAPTVSVTVTPGKPLMTVTTLKRPSSLGCASTPSENPDLKAENSA	180
Oy	181	VOINISPTMLEVKKCKNFMLMLIKLACSSQSPMGQNNKLYEQLLDAKIEAEETFRK	240
Db	181	VOINISPTMLEVKKCKNFMLMLIKLACSSQSPMGQNNKLYEQLLDAKIEAEETFRK	240
Oy	241	LYVELKSSPOPILVPEFLKKSVAVALRQLLPNSOSFIQOCVVOQTSSDMVATCTTTVTTS	300
Db	241	LYVELKSSPOPILVPEFLKKSVAVALRQLLPNSOSFIQOCVVOQTSSDMVATCTTTVTTS	300
Oy	301	VTVTVSSSSQSEKSIIVSGATAPRTVSVOTLNLPLAGPYGAKAGVTLNLSVGPPTAATGCTTA	360
Db	301	VTVTVSSSSQSEKSIIVSGATAPRTVSVOTLNLPLAGPYGAKAGVTLNLSVGPPTAATGCTTA	360
Oy	361	GTGLIQTSPKPLVTSVAANTVTTVTSLOPEKRPVVSQGNATVLSLPAVTGSETSGAICLP	420
Db	361	GTGLIQTSPKPLVTSVAANTVTTVTSLOPEKRPVVSQGNATVLSLPAVTGSETSGAICLP	420
Oy	421	VVSPFMDHICKPVIGTPVOIKLAGPVLQSP	452
Db	421	VVSPFMDHICKPVIGTPVOIKLAGPVLQSP	452
RESULT 4			
AAR56494			
AC	XX	AAR56494 standard; Protein: 737 AA.	
AC	XX	AAR56494;	
DF	XX	23-MAR-1995 (first entry)	
DE	XX	TARA-binding protein-associated factor hTAFII130.	
DE	XX	TARA-binding protein-associated factor; hTAFII130; screening;	
KM	XX	diagnostic; therapeutic; gene transcription regulation.	
OS	XX	Homo sapiens.	
PN	XX	W09417087-A.	
PD	XX	04-AUG-1994.	
PF	XX	28-JAN-1994; 94MO-US01114.	
PR	XX	28-JAN-1993; 93US-0013412.	
PR	XX	30-JUN-1993; 93US-0087119.	
PA	XX	(REGC) UNIV CALIFORNIA.	
PI	XX	Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;	
PI	XX	Tjian R, Wang E, Weinzierl ROJ;	
DR	XX	WPI: 1994-264019/32.	
DR	XX	N-PSDB: AAO70731.	
PT	XX	TARA-binding protein associated protein factors - and	
PT	XX	corresponding nucleotide sequence and deriv. antibodies, useful	
PS	XX	in screening, diagnostics and therapeutics	
PS	XX	Disclosure: Page 142; 180pp; English.	
CC	XX	The TARA-binding protein associated factor hTAFII130 (including	
CC	XX	specific antibodies and fusion products) are used in drug screening,	
CC	XX	diagnostics and therapeutics. They are used in the development of	
CC	XX	specific biochemical assays for screening compounds that agonise or	
CC	XX	antagonise selected transcription factors involved in regulating	
CC	XX	gene expression associated with human pathology.	
SO	XX	Sequence 737 AA;	

Query Match 23.2%; Score 521; DB 15; Length 737;
 Best Local Similarity 34.0%; Pred. No. 3.2e-30;
 Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

13 PKVSSG-----PRLPAPQIVAVKAPMTTITQFPANILQPPGVILIKNSGPILM 61
 68 PRATTSIGIRATLMPVLAAPRLPQ-----PQNPNTNQ-----NQLPFGMWLVSENGQL 118
 62 LVSPOQTVTR-----AETTSNITSRPAPVAPNPQVYKICVNPSSSOLIKKVAATPYKYL 115
 119 MI-POQALAQMOQAHAQOQTWAPRPAPPTSAAPVOISTVQAGPPIIAR-CVTP----- 172
 116 AQTGTVVTVPKSSVQSAVPTSVTVTPGKPLNVT--TLKPSLIGSS-----TPS 168
 173 ----TTIIKQV---SQAQTVPQSAITQNSPGVQPOLVLGGAQOTSLGTAIVQGTGPQ 225
 169 NE-PNLKASNAVQINLSPTMLENVKCKNFAMLIKILACSGSQSPSEMONGYKLIQEO 227
 226 RVVPGATTTSSATE-----TMEVVKCKNFSTLIKILACSGSQSTETANVKELVQNL 279
 228 LDKAIAEETRLKLYELKSSPOPHLVPLKKSVALROLPPNSQSTIQOCVQO---TS 283
 280 LDKAIAEETRLKLYELKSSPOPHLVPLKKSVALROLPPNSQSTIQOCVQO---TS 339
 284 SDWVIATCTTGTVPKSSVQSAVPTSVTVTPGKPLNVT--TLKPSLIGSS----- 343
 340 Q-----ATTALITAVLVSSVQRTKATATVTSALQPPVLSL----- 376
 344 VTLHSVGPATAGTGTAGTGLQTSKPLVTSVANTVTVTSALQPEKPVSGTAVTSLP 403
 377 -----TQPTGVGKQOGPPPLVQ-----GPPKP-----GALINRPQV 410
 404 TEGETSGAALCLPSKPVVSVFCDHICKPVYIGPVQIKL--AOPGPVL 449
 411 TLTQT-----PMVALRQPH-NRIMLTTPQOIPLPQVPPV 446

RESULT 5
 AAM06084
 ID AAM06084 standard; Protein: 737 AA.
 XX AAM06084;
 AC AAM06084;
 XX 27-JAN-1997 (first entry)
 DT 27-JAN-1997 (first entry)
 XX Human TATA-binding protein associated factor hTAFII130 protein.
 DE Human TATA-binding protein associated factor hTAFII130 protein.
 XX Human; TATA-binding protein; TBP associated factor; TFIID; holoenzyme;
 KM RNA polymerase II; transcription; messenger RNA; nuclear fraction;
 KM lambda-gli1; expression library.
 XX Homo sapiens.
 OS Homo sapiens.
 XX US5534410-A.
 PN US5534410-A.
 XX 09-JUL-1996.
 PD 09-JUL-1996.
 XX 28-JAN-1993; 93US-0013412.
 PF 28-JAN-1993; 93US-0013412.
 XX 28-JAN-1994; 94US-0188582.
 PR 28-JAN-1993; 93US-0013412.
 PR 30-JUN-1993; 93US-0087119.
 XX (REGC) UNIV CALIFORNIA.
 PA (REGC) UNIV CALIFORNIA.
 XX Comal L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
 PI Tjian R, Wang E, Weinzierl RJ;
 XX WPT: 1996-333245/33.
 DR N-PSDB: AAT42217.
 XX Screen for cpts. that bind human TATA-binding protein associated

PT factor - by testing ability to bind to polypeptide fragments of the
 PT factor, useful as (ant)agonists of transcription factors involved in
 PT disease.
 XX Examples: Column 105-112; 86pp; English.
 PS This is the amino acid sequence of the human TATA-binding protein (TBP)
 CC associated factor (TAF) designated TAFII130. The protein is a component
 CC of the TFIID fraction required for reconstituting RNA polymerase II in
 CC vitro transcription activity. The encoded protein has an estimated mol.
 CC wt. of 130 kD by SDS-PAGE.
 CC The invention relates to purified proteins involved in transcription
 CC by RNA polymerase II, the RNA polymerase which transcribes messenger
 CC RNA. RNA polymerase II transcription proceeds in vitro upon addition
 CC of several nuclear fractions designated TFIID, B, D, E, F, H, I and J
 CC to RNA polymerase II holoenzyme. Fraction TFIID has been shown to
 CC contain a TBP and other TAFs. Purification of TFIID and separation of
 CC its components reveals 7 proteins ranging in size from 30-250 kD.
 CC Serum raised against the TFIID fraction allowed cloning of the corresp.
 CC genes from lambda-gli1 expression libraries.
 XX Sequence 737 AA:

Query Match 23.2%; Score 521; DB 17; Length 737;
 Best Local Similarity 34.0%; Pred. No. 3.2e-30;
 Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

13 PKVSSG-----PRLPAPQIVAVKAPMTTITQFPANILQPPGVILIKNSGPILM 61
 68 PRATTSIGIRATLMPVLAAPRLPQ-----PQNPNTNQ-----NQLPFGMWLVSENGQL 118
 62 LVSPOQTVTR-----AETTSNITSRPAPVAPNPQVYKICVNPSSSOLIKKVAATPYKYL 115
 119 MI-POQALAQMOQAHAQOQTWAPRPAPPTSAAPVOISTVQAGPPIIAR-CVTP----- 172
 116 AQTGTVVTVPKSSVQSAVPTSVTVTPGKPLNVT--TLKPSLIGSS-----TPS 168
 173 ----TTIIKQV---SQAQTVPQSAITQNSPGVQPOLVLGGAQOTSLGTAIVQGTGPQ 225
 169 NE-PNLKASNAVQINLSPTMLENVKCKNFAMLIKILACSGSQSPSEMONGYKLIQEO 227
 226 RVVPGATTTSSATE-----TMEVVKCKNFSTLIKILACSGSQSTETANVKELVQNL 279
 228 LDKAIAEETRLKLYELKSSPOPHLVPLKKSVALROLPPNSQSTIQOCVQO---TS 283
 280 LDKAIAEETRLKLYELKSSPOPHLVPLKKSVALROLPPNSQSTIQOCVQO---TS 339
 284 SDWVIATCTTGTVPKSSVQSAVPTSVTVTPGKPLNVT--TLKPSLIGSS----- 343
 340 Q-----ATTALITAVLVSSVQRTKATATVTSALQPPVLSL----- 376
 344 VTLHSVGPATAGTGTAGTGLQTSKPLVTSVANTVTVTSALQPEKPVSGTAVTSLP 403
 377 -----TQPTGVGKQOGPPPLVQ-----GPPKP-----GALINRPQV 410
 404 TEGETSGAALCLPSKPVVSVFCDHICKPVYIGPVQIKL--AOPGPVL 449
 411 TLTQT-----PMVALRQPH-NRIMLTTPQOIPLPQVPPV 446

RESULT 6
 AAM25019
 ID AAM25019 standard; Protein: 737 AA.
 XX AAM25019;
 AC AAM25019;
 XX 08-OCT-1997 (first entry)
 DT 08-OCT-1997 (first entry)
 XX TATA-binding protein associated factor, hTAFII130.
 DE TATA-binding protein associated factor, TAF; nuclear protein;
 KM RNA polymerase II; transcription; TATA-binding protein; TBP;
 KM RNA polymerase II; transcription; TATA-binding protein; TBP;
 KM Initiation.


```
XX OS Homo sapiens.
XX PN US637686-A.
XX PD 10-JUN-1997.
XX PE 28-JAN-1993; 93US-0013412.
XX PR 28-JAN-1994; 94US-0188582.
XX PR 28-JAN-1993; 93US-0013412.
XX PR 30-JUN-1993; 93US-0087119.
XX PR 09-MAY-1996; 96US-0646715.
XX PA (RECC ) UNIV CALIFORNIA.
XX PI Comal L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
XX PI Tjian R, Wang E, Weinzierl ROJ;
XX DR WPI: 1997-319113/29.
XX DR N-PSDB; AAT79595.
XX PT Nucleic acids encoding human TATA-binding protein associated factor
XX PT (TAF) peptide(s) - for production of recombinant peptide(s), used
XX PT for modulating transcription of TAFs
XX PS Claim 1; Column 111-116; 86pp; English.
XX CC AAW25018 represents TATA-binding protein associated factor (TAF)
XX CC polypeptide, hTAFII100 (mol. weight 100kd). TAF peptides derived
XX CC from hTAFII100 alpha, hTAFII100 beta, hTAFII100, hTAFII100,
XX CC hTAFII130, hTAFII250, hTAFII48 and hTAFII10 and nucleic acids encoding
XX CC them, are used to modulate transcription, including transcription
XX CC initiation. TAFs are nuclear proteins involved in RNA polymerase I,
XX CC II and III transcription. The peptides act by binding to a different
XX CC TAF, an activator, or TBP (TATA-binding protein) or competitively
XX CC inhibiting association of a TAF domain with another compound, typically
XX CC a protein like TBP or another TAF, an activator, or DNA.
XX SQ Sequence 737 AA;
XX
XX Query Match 23.2%; Score 521; DB 18; Length 737;
XX Best Local Similarity 34.0%; Pred. No. 3.2e-30;
XX Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;
XX
XX 13 PPKVSSG-----PRLPAQIVAVKAPNTTITQFPANLQIPGVILIKNSGPILM 61
XX | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 68 PTATTSGRATLTPTVLAIRLPDP-----PQNTPTNIO--NFQLPFGMVLVRSNGQLL 118
XX | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX QY 62 LVSPQGVYTR-----AETTSNITSRPVAPNPOTVKICTVPSNSSQILIKKVAVTPVKRL 115
XX | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX DB 119 MI-PQQAIAQMAQAHAPQPTTMARPRAPTPTSPAPVQISTVOAGCTPIIAR-QVTP----- 172
XX | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX QY 116 AQIGTGVVTVTPKPSVQSAVNPVSIVTTPGKPLNTYT--TLKPSISGASS-----TPS 168
XX | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX DB 173 ----TIIKQV---SQAGTVQPSATLQSRSPGVQQLVLGGAAGVASTGTAFAVQGTGPQ 225
XX | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX QY 169 NE-PNLKENSAAVQINLSPTMLENVKCKNFILMLIKIACSGSPGMCQNVKVLVQQL 227
XX | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX DB 226 RTVGATTTSSAATE-----TMENVKCKNFILSLTIKASSGOSTTANAVKELVQNL 279
XX | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX QY 228 LDKATIEAEFRKRLKYLEKSSPQHLVDFLKSSVALKQLLPNSOSFTQCGVQO----TS 283
XX | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX DB 280 LDGKIEADEDFSRILYRELNNSPOPYLVFELKRSIPALQLPDPDSAAFIQSQOQPPPTPS 339
XX | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX QY 284 SDWVIATCTTVTSPVYTTVSSSQSEKSTIVSGATAPRVSVQTLNPLAGPVAKAGV 343
XX | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX DB 340 Q-----ATTAALAAVLAASSVQRTAGCTAATVTSALQPVLSL----- 376
XX | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX QY 344 VTLHSGPFAATGCTAGTGLQTSKPLVTSVANTVTTVSLQDEKPVVSGTAVTLSPAV 403
XX | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX DB 377 -----TQPTQVGVGKGGQPTPLVIG-----QPPRP-----GALIRPQV 410
XX | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
OY 404 TFGETSGAAILCPYKVPVSCWDHICKPVIGTPVOIKL--AQPGPVL 449
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 411 TLTFQF-----PMVALRQPH-NRIMLTTPQOIQLNPLQPVPPV 446
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX
XX RESULT 7
XX ID AA082954 standard; Protein; 1023 AA.
XX AC AA082954;
XX XX
XX 23-APR-2002 (first entry)
XX DE
XX DE Human homologue of MP11 protein target for antifungal compound.
XX KW
XX KW Antifungal: fungal gene transcription; RPC34; POP3; TFA2; NAB2;
XX KW MPT1; MTR2; BOS1; POL30; RSA2; SGT1; MTR1; TFB1; SPC98; BFR2; RNA1;
XX KW GCD7; SK16; NIP1; LCP5; NCE103; ECO1; ORC2; CNS1; YPD1; TIM10; SRB4;
XX KW yeast; fungus.
XX OS
XX OS Homo sapiens.
XX PN WO20020205-A2.
XX PD 10-JAN-2002.
XX PE 28-JUN-2001; 2001WO-US20592.
XX PF 29-JUN-2000; 2000US-215164P.
XX PR 10-AUG-2000; 2000US-224457P.
XX PA (ANAD-) ANADYS PHARM INC.
XX PI Moore J, Buurman ET, Desilva T, Harris S, Komarnitsky S;
XX PI Mendillo M, Moore D, McCoy M, Sanderson K, Haq T, Zhu S, Long F;
XX PI Davidov E, Thompson CM;
XX DR WPI: 2002-147962/19.
XX DR N-PSDB; ABK32842.
XX
XX PT Screening candidate antifungal compound for interaction with essential
XX PT protein, modulation of essential protein activity, binding to essential
XX PT protein, by contacting protein with test compound and determining
XX PT effects
XX PS Claim 1; Figure 79; 522pp; English.
XX
XX CC The invention describes a method of screening a candidate antifungal
XX CC compound for interaction with essential proteins (EP) or for modulation
XX CC of EP activity e.g fungal gene transcription. The proteins tested in the
XX CC invention include RPC34, POP3, TFA2, NAB2, MPT1, MTR2, BOS1, POL30, RSA2,
XX CC SGT1, MTR1, TFB1, SPC98, BFR2, RNA1, GCD7, SK16, NIP1, LCP5, NCE103,
XX CC ECO1, ORC2, CNS1, YPD1, TIM10 and SRB4 from S. cerevisiae, C. albicans
XX CC and human homologues. The method involves contacting a culture with one
XX CC or more test compounds and determining the effects on the growth or
XX CC viability of the culture of cells which preferably comprises fungal cells
XX CC or yeast cells. Preferably the identified compounds interact with, or
XX CC modulate (preferably inhibit) activity of C. albicans EP. The inhibitor
XX CC compounds identified by the method are useful for preventing or
XX CC inhibiting fungal, particularly C. albicans growth in culture or in a
XX CC mammal. The antifungal agents interact with essential fungal elements
XX CC that can be used to treat fungal infection by preventing the growth and
XX CC preferentially killing the fungi, but does not inhibit the biological
XX CC activity of mammalian homologues. This amino acid sequence represents a
XX CC target protein used to test the antifungal compounds, described in the
XX CC method of the invention.
XX SQ Sequence 1023 AA;
XX
XX Query Match 22.8%; Score 513; DB 23; Length 1023;
XX Best Local Similarity 33.4%; Pred. No. 2e-29;
XX Matches 154; Conservative 62; Mismatches 137; Indels 118; Gaps 19;
```

```

QY 13 PKVSSG-----PRLPAPQIVAVKAPNTTIOFPANIOIPGCTVILKNSGRLM 61
DB 414 PNTTISRATLPTTUALPRLPQ-----PQNTNIO-----NEFLPGGMVLVRENGQL 464
QY 62 LVSPQQTIVR-----AETTSNITSRPVAPNPQTVKICITVPSNSOLIKKVAATPVKTL 115
DB 465 MI-PQQAQAQMAQAHQAPQTTMAPRAPATPTSAPVQISIVQAGCTIILAR-QVTP----- 518
QY 116 AQGTIVTVTVTPKPSVQSAVAPTSVTVTPGKPLNTVT--TLKPSGLASS-----TPS 168
DB 519 -----TTIKQV-----SQAQTVQPSATLQSRPGVOPQVLGGAQGTASLGTATVQGTGPQ 571
QY 169 NE-PNLKAENSAVQIUNLSPMLNENVKCKNPLAMLIKLAOSGSPSEMGONVKLVEOL 227
DB 572 RTYPGATTTSSAATE-----TMENNAKCKNFKSTLIKLAASSGKQSTETANNAVELQNL 625
QY 228 LDARKEAEETRLKLYELKSSPOPHLVPFLKSVVALROLPLNSQSTFOOCVQO-----TS 283
DB 626 LDGKIEADEPTSLRRELNSPQYLVPLKRSLLPALROLPLDSEANFIQSSQOQPPPTIS 685
QY 284 SDAVIATCTTIVTSPVVTTVSSOSEKSIYSGATAPRTVSQIUNPLAGPVGAKAGV 343
DB 686 Q-----ATTALTAAVVLSSSVQRTAGKTATATYSALQPPVLSL----- 722
QY 344 VTLHSVGPPLAATGTTAGTGLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAATLSLPAV 403
DB 723 -----TOPTOYGVCKQGOPTPLVIO-----QPPK-----GALIRPPQV 756
QY 404 TFGETSGAMICLPSKPVVSCFMDHICKPVIGTPVQIKLAQ 444
DB 757 TLEQT-----PMVALRQPH-NRIMLTTPQOVNLSE 785

RESULT 8
AAR56487
ID AAR56487 standard; Protein; 921 AA.
AC AAR56487;
XX
XX 23-MAR-1995 (first entry)
XX
XX TATA-binding protein-associated factor dtraf1110.
DE
XX TATA-binding protein associated factor; dtraf1110; screening;
KM diagnostic; therapeutic; gene transcription regulation.
XX
XX Drosophila.
OS
XX WO9417087-A.
PN
XX 04-AUG-1994.
PD
XX 28-JAN-1994; 94WO-US01114.
PE
XX 28-JAN-1993; 93US-0013412.
PR 30-JUN-1993; 93US-0087119.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Comal L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
PI Tjian R, Wang E, Weinzierl ROJ;
XX
XX MPI: 1994-264019/32.
DR N-PSDB: AAO70724.
XX
XX TATA-binding protein associated protein factors - and
PT corresponding nucleotide sequence and deriv. antibodies, useful
XX in screening, diagnostics and therapeutics
XX
XX Disclosure: Page 56-61; 180pp; English.
PS
XX
XX The TATA-binding protein associated factor dtraf1110 (including
CC specific antibodies and fusion products) are used in drug screening,

```

```

CC diagnostics and therapeutics. They are used in the development of
CC specific biochemical assays for screening compounds that agonise or
CC antagonise selected transcription factors involved in regulating
CC gene expression associated with human pathology.
XX
XX Sequence 921 AA;
XX
XX Query Match 10.7%; Score 241.5; DB 15; Length 921;
XX Best Local Similarly 24.8%; Pred. No. 2.8e-09;
XX Matches 112; Conservative 65; Mismatches 163; Indels 111; Gaps 17;
XX
QY 25 POIYAVKAPNTTIOFPANIOIPGCTVILKNSGRLM-----VSPQQTIVRAETTS 76
DB 133 PQSPSTLSLNLNTGQTPA-----LLVKTNDGFOILRVGTTGPPVTOTITNSNS 184
QY 77 NITSRPVAPNPQTVKICITVPSNSQ-----LIKVAVTPVKRLAIGTVTVTPV 127
DB 185 NTSTTNHPTTQ-IRLOTVPAASMTNTATSNIIYNSVAASSGVANSSQPHILTOINAQ 243
QY 128 KPSVQSAVAPTSVTVTPGKPLNTVTTLKPSGLASSFTSNEPNLKAENSAVQIUNLSP 187
DB 244 APOLPQITQIOTIPAAQSOQOQVNNVSSAGGTATVAVSSITA-----ATG 287
QY 188 TMLNENVK-KCKNPLAMLIKLAOSGSPSEMGONVKLVEOLDAKIEAEETRLKLYELK 246
DB 288 TQCGNTKECKKFLANIEL--STREPKPVKKNVRLIOELVANVBEFCDLRLN 345
QY 247 SSPQHLVPLFKSVVALROL-----LPNSQSTFO----- 276
DB 346 ASPQPCILGFLKSLPLRQALYKELVIGIKRPPQHVIGIAGLSQDLKIKQAQITPIC 405
QY 277 ----QCVOQTSMDNVATCTTIVTSPVVTTVSSOSEKSIYSGATAPRTVS--VQT 329
DB 406 PSQTTITIGQTOVRMI--TPNALGTPRPTIGTTISKQPN--IRLEPTADRLVNTGIRT 460
QY 330 LNPAGPVGAKAGVTVLHSVGPPLAATGTTAGTGLQTSKPLVTSVANTVTVTSLOPEK 389
DB 461 QIP-SLOVPGQANIVQIR--GPOHAQLQRTGSVOIRATRTP-----PNSVPTAN----- 506
QY 390 VVSGTAATLSLPAVTFGETSGAMICLPSVVP 420
DB 507 -----KLTVAVKVGQTOIKAI-TPSLMP 527

RESULT 9
AAW06077
ID AAW06077 standard; Protein; 921 AA.
AC AAW06077;
XX
XX 27-JAN-1997 (first entry)
XX
XX Drosophila TATA-binding protein associated factor dtraf1110 protein.
DE
XX Drosophila; TATA-binding protein; TBP associated factor; TFIID;
KM RNA polymerase II; transcription; messenger RNA; nuclear fraction;
XX holoenzyme; lambda-gli1; expression library.
XX
XX Drosophila melanogaster.
OS
XX US5534410-A.
PN
XX 09-JUL-1996.
PD
XX 28-JAN-1993; 93US-0013412.
PE
XX 28-JAN-1994; 94US-0186582.
PR 28-JAN-1993; 93US-0013412.
XX 30-JUN-1993; 93US-0087119.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Comal L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
PI

```

PI Tjian R, Wang E, Weinzierl ROJ;
 XX WPI: 1996-333245/33.
 DR N-PSDB; AAT42210.
 XX
 PT Screen for cpds. that bind human TATA-binding protein associated
 PT factor - by testing ability to bind to polypeptide fragments of the
 PT factor, useful as (ant)agonists of transcription factors involved in
 PT disease.
 PS
 XX Examples: Column 27-36; 86pp; English.
 CC This is the amino acid sequence of the Drosophila TATA-binding protein
 CC (TBP) associated factor (TAF) designated TAF110. The protein is a
 CC component of the TFIID fraction required for reconstituting RNA
 CC polymerase II in vitro transcription activity. The encoded protein
 CC has an estimated mol. wt. of 110 kD by SDS-PAGE, a calculated mol. wt.
 CC based on sequence of the gene of 99.4 kD and an estimated pI of 10.1.
 CC The invention relates to purified proteins involved in transcription
 CC by RNA polymerase II, the RNA polymerase which transcribes messenger
 CC RNA. RNA polymerase II transcription proceeds in vitro upon addition
 CC of several nuclear fractions designated TFIIA, B, D, E, F, H, I and J
 CC to RNA polymerase II holoenzyme. Fraction TFIID has been shown to
 CC contain a TBP and other TAFs. Purification of TFIID and separation of
 CC its components reveals 7 proteins ranging in size from 30-250 kD.
 CC Serum raised against the TFIID fraction allowed cloning of the corresp.
 CC genes from lambda-gt11 expression libraries.
 XX
 XX Sequence 921 AA;
 SO
 Query Match 10.7%; Score 241.5; DB 17; Length 921;
 Best Local Similarity 24.8%; Pred. No. 2.8e-09;
 Matches 112; Conservative 65; Mismatches 163; Indels 111; Gaps 17;
 QY 25 POIYAVKAPNTTTIOFPANLQLPGETVLKISNSGPLM-----VSPQOTVRAETTS 76
 DB 133 POSSITLSTANTGCTPA-----LLVKTDNGFQLRVTGTPPTVOTINTSNNS 184
 QY 77 NITSRPAPNPQTVKICTVPSNSSQ-----LKKVAVTVKKLAQIGTVVTVTP 127
 DB 185 NITSTNHPPTTQ-IRLOTVPAAASMTNTATSNINVASSGVANSQPHLTLQNAQ 243
 QY 128 KPSSQVAVPFTSVYTVYVPGKPLNTVTLKPSLGASTPSENEPMLKENSAAVOINLSP 187
 DB 244 APOLPQIOTIOTIPAOQSOQOVNNSAGCTATAVSSSTA-----ATT 287
 QY 188 TMLBNVK-KCKNFLAMLIKLAGSGSOSPDMGQNVKKEVLLDKIAEEFTRLIYELK 246
 DB 288 TQOGNTEKCKRFLANLIEL--STREPKPYEKNVRLTIQELVANNVBEFECDRLERLIN 345
 QY 247 SSPQHLVPLPKSSVALROL-----LPSQSFTQ--- 276
 DB 346 ASPQCLIGLKSLPLRLQALYRKELVIEGIRPPQHVGLAGLSQQLPQIQAIRPIG 405
 QY 277 -----QCVQOTSMDVIATCTTTTSPYVTTVSSSSEKSIIVSGATAPRTVS---VQT 329
 DB 406 PSQTTTIGTQOVNRM--TPNALGTPRPPIGHTTISKQPPN--IRLTPAPRLVWVGIRT 460
 QY 330 LNPAGPVGAKAGVVTLSHVGPTAATGCTAGTGLTOSPLVTSVANTVTVLSQPKP 389
 DB 461 QIP-SLOYPGGANIVQIR--GPOHAQLOKRTSSVOIRATIRP-----PNSVPTAN----- 506
 QY 390 VVSGTAATLTLPAVTFGTSGAALCPSPVKP 420
 DB 507 -----KLTAVKVGOTQIKAL-TPSLHP 527
 RESULT 10
 ID AAM25028 standard; protein; 921 AA.
 XX
 AC AAM25028;
 XX

DT 08-OCT-1997 (first entry)
 XX
 DE TATA-binding protein associated factor, dTAF110.
 XX
 XX TATA-binding protein associated factor; TAF; nuclear protein;
 KW RNA polymerase transcription; TATA-binding protein; TBP;
 KW initiation.
 XX
 OS Drosophila sp.
 XX
 PN US5637686-A.
 XX
 PD 10-JUN-1997.
 XX
 XX 28-JAN-1993; 93US-0013412.
 XX
 PR 28-JAN-1994; 94US-0188592.
 PR 28-JAN-1993; 93US-0013412.
 PR 30-JUN-1993; 93US-0087119.
 PR 09-MAY-1996; 96US-0646715.
 XX
 PA (REGC) UNIV CALIFORNTA.
 XX
 PI Comal L, Dynlact BD, Hoey T, Rupert S, Tanese N;
 PI Tjian R, Wang E, Weinzierl ROJ;
 XX
 DR WPI: 1997-319113/29.
 DR N-PSDB; AAT79604.
 XX
 XX Nucleic acids encoding human TATA-binding protein associated factor
 PT (TAF) peptide(s) - for production of recombinant peptide(s), used
 PT for modulating transcription of TAFs
 XX
 PS Example 1; Column 35-40; 86pp; English.
 XX
 XX AAM25028 represents TATA-binding protein associated factor (TAF)
 CC polypeptide, dTAF110 (mol. weight 110kD). TAF peptides derived
 CC from dTAF110 alpha, dTAF110 beta, dTAF110, dTAF110, dTAF110,
 CC dTAF110, dTAF110, and dTAF110, their human equivalents and
 CC nucleic acids encoding them, are used to modulate transcription,
 CC including transcription initiation. TAFs are nuclear proteins involved
 CC in RNA polymerase I, II and III transcription. The peptides act by
 CC binding to a different TAF, an activator, or TBP (TATA-binding protein)
 CC or competitively inhibiting association of a TAF domain with another
 CC compound, typically a protein like TBP or another TAF, an activator,
 CC or DNA.
 CC
 SO Sequence 921 AA;
 Query Match 10.7%; Score 241.5; DB 18; Length 921;
 Best Local Similarity 24.8%; Pred. No. 2.8e-09;
 Matches 112; Conservative 65; Mismatches 163; Indels 111; Gaps 17;
 QY 25 POIYAVKAPNTTTIOFPANLQLPGETVLKISNSGPLM-----VSPQOTVRAETTS 76
 DB 133 POSSITLSTANTGCTPA-----LLVKTDNGFQLRVTGTPPTVOTINTSNNS 184
 QY 77 NITSRPAPNPQTVKICTVPSNSSQ-----LKKVAVTVKKLAQIGTVVTVTP 127
 DB 185 NITSTNHPPTTQ-IRLOTVPAAASMTNTATSNINVASSGVANSQPHLTLQNAQ 243
 QY 128 KPSSQVAVPFTSVYTVYVPGKPLNTVTLKPSLGASTPSENEPMLKENSAAVOINLSP 187
 DB 244 APOLPQIOTIOTIPAOQSOQOVNNSAGCTATAVSSSTA-----ATT 287
 QY 188 TMLBNVK-KCKNFLAMLIKLAGSGSOSPDMGQNVKKEVLLDKIAEEFTRLIYELK 246
 DB 288 TQOGNTEKCKRFLANLIEL--STREPKPYEKNVRLTIQELVANNVBEFECDRLERLIN 345
 QY 247 SSPQHLVPLPKSSVALROL-----LPSQSFTQ--- 276
 DB 346 ASPQCLIGLKSLPLRLQALYRKELVIEGIRPPQHVGLAGLSQQLPQIQAIRPIG 405

QY 277 ----OCVOQTSSDMVIATCTTIVTTSPPVYTVTVSSSSOSEKSIIVSGATAPRTVS---VQT 329
DB 406 PSQTTTIGQTOVAMI--TPNALGTPRTTIGHTTISKQPPN---IRLPTARLVNTGIGIRT 460
QY 330 LNPAGPVGAKAGVNTLHSGVPTAATGTTAGTGLQTSKPLVTSVANTVTVSLOPEKP 389
DB 461 QIP-SLQVPGQANIVQIR--GPQHAQLQRTGVSQIRATTRP-----PNSVPTAN----- 506
QY 390 VSGTAVTISLPAVTFGETSGAICLPVSKP 420
DB 507 -----KLTAVKVGQTOIKAI-TPSLHP 527

RESULT 11
ABB61528
ID ABB61528 standard; Protein; 921 AA.
AC ABB61528;
XX
XX
XX 26-MAR-2002 (first entry)
DT
DE Drosophila melanogaster polypeptide SEQ ID NO 11376.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical.
XX
XX Drosophila melanogaster.
OS
PN WO200171042-A2.
XX
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA
PI Venter JC, Adams M, Li PMD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR N-PSDB; ABL05631.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX
PS Disclosure; SEQ ID NO 11376; 21np + Sequence Listing; English.
XX
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
CC sequences (ABLI01840-ABLI6175) and the encoded proteins
CC (ABBS7737-ABBS72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 921 AA:

Query Match 10.7%; Score 241.5; DB 22; Length 921;
Best Local Similarity 24.8%; Pred. No. 2.8e-09;
Matches 112; Conservative 65; Mismatches 163; Indels 111; Gaps 17;

QY 25 PQIVAKAPNTTITQPPANQLPFGIVLIKNSGPMLE-----VSPOQVTRAEETS 76
DB 133 POSPSTLSTLNTGQTPA-----LWKTDNGFOLLRLRGTTGPPVQQTITNTSNN 184
QY 77 NITSRAVAVANPQTVKICVTPNSSQ-----LIKVAVTPVKKLAQIGTIVTVTP 127

DB 185 NTSTTNHPPTTO-IRLQIVPAASMTNTTATNTIIVNSVASSGIANSSOPPHLTOLMAO 243
QY 128 KPSSVQSAVPTSVTVTPCKPLNTVYTLKPSISIGASPTSPNBPNUKAENSAVQINLSP 187
DB 244 APQLPQITQITIPAQOSQOQVNNVSSAGCTADAVSSTA-----ATT 287
QY 188 TMLBNK-KCKNFLAMIKIACSGSOSPEMGQVKKVLEDLLDAKIEAEFTKKLYELK 246
DB 288 TQOGNTEKCKRFLANLIEI--STRBKPVEKKNVRLIOLVNAVNEPEFCRLRLN 345
QY 247 SSPQHLVPELKKSVVALRQL-----LPSQSFIQ--- 276
DB 346 ASPQPLIGFLKSLPRLRALTKELYEGIRPPQHVGLAGLSQQLKIQALRPIG 405
QY 277 ----OCVOQTSSDMVIATCTTIVTTSPPVYTVTVSSSSOSEKSIIVSGATAPRTVS---VQT 329
DB 406 PSQTTTIGQTOVAMI--TPNALGTPRTTIGHTTISKQPPN---IRLPTARLVNTGIGIRT 460
QY 330 LNPAGPVGAKAGVNTLHSGVPTAATGTTAGTGLQTSKPLVTSVANTVTVSLOPEKP 389
DB 461 QIP-SLQVPGQANIVQIR--GPQHAQLQRTGVSQIRATTRP-----PNSVPTAN----- 506
QY 390 VSGTAVTISLPAVTFGETSGAICLPVSKP 420
DB 507 -----KLTAVKVGQTOIKAI-TPSLHP 527

RESULT 12
ABB66055
ID ABB66055 standard; Protein; 921 AA.
AC ABB66055;
XX
XX
XX 26-MAR-2002 (first entry)
DT
DE Drosophila melanogaster polypeptide SEQ ID NO 24957.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical.
XX
XX Drosophila melanogaster.
OS
PN WO200171042-A2.
XX
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA
PI Venter JC, Adams M, Li PMD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR N-PSDB; ABL10158.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX
PS Disclosure; SEQ ID NO 24957; 21np + Sequence Listing; English.
XX
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
CC sequences (ABLI01840-ABLI6175) and the encoded proteins
CC (ABBS7737-ABBS72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 921 AA:

Query Match 10.78; Score 241.5; DB 22; Length 921;
Best Local Similarity 24.84; Pred. No. 2.8e-09;
Matches 112; Conservative 65; Mismatches 163; Indels 111; Gaps 17;

QY 25 POIYAVKAPMTTPTIOPANQLPGTVLINSNGPLM-----VSPQOTVRAEFTS 76
DB 133 POSSITLSTLNTGQTA-----LVKTDNSFQLRGTGTPPTVOTITNTSNN 184
QY 77 NITSPAVPANPQTVKICTVPNSSQ-----LIRKVAVPVKIKLAQIGTVVTVTP 127
DB 185 NITSTTNHPTTQ-IRLOTVPAASMTNTATSNIIYNSVASSGYANSSQPHLTQNAQ 243
QY 128 KPSSVGVAVPTSVYVTPKPLNTVTTTLKPSLSGASTSNEPNLKAENSAVAQINLSP 187
DB 244 APOLPOTIOTIIPAOOSOOOVNNSAGGTATAVSSTA-----ATT 287
QY 188 TMLENVK-KCKNPLAMLIKACSGSPENGQNKLVLEOLDKIRAEFTRKLYELK 246
DB 288 TQCGNTEKCKRPLANIEL--STREPKVEKNVRILOELVANNVEPEECDFLERLIN 345
QY 247 SSPQPHLVPFLKKSVALROL-----LPSQSFIO--- 276
DB 346 ASPQCLIGFLKSLPRLQALYKELVIGIKPPQHVGLAGLSQQLPQIQAIRPIS 405
QY 277 ----QCVOQSSDMVATCTTVTTSVVTTSVSSQSEKSIYSGATAPRTVS---VQT 329
DB 406 PSQTTTIGQOVNMI--TPNALGTPRPTIGTITISKOPN--IRLPAPRLVMTGIR 460
QY 330 LNPAGVAGKAGVPTLHSGPTAATGTTAGTGLDTSKPLVTSVANTVTVSLQPEK 389
DB 461 QIP-SLVYPCQANIVQIR--GPOHQAQORTGSVOIRATRP-----PNSVPTAN----- 506
QY 390 VSGTAVTSLPAVTFCTSGAICLPSPVK 420
DB 507 -----KLTAVKVGOTQIKAI-TPSLHP 527

RESULT 13
AAR57141
ID AAR57141 standard; Protein; 2035 AA.

XX AAR57141;

DT 19-MAR-1995 (first entry)

DE Host cell factor protein.

KM Herpes simplex virus; herpes virus; VP16; immediate early gene;
KW host cell factor; virus infection therapy; cellular protein;
transcription.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Active-site 21..31

FT Active-site /note="peptide R60"

FT Active-site 168..186

FT Active-site /note="peptide R37"

FT Active-site 333..340

FT Active-site /note="peptide R52"

FT Active-site 426..449

FT Active-site /note="peptide 362"

FT Active-site 511..526

FT Active-site /note="peptide 329"

FT Active-site 578..594

FT Active-site /note="peptide 223 first sequence"

FT Active-site 594..611

FT /note="peptide R26 1st peptide"
FT 611..623
FT /note="peptide 223 2nd sequence"
FT 723..731
FT /note="peptide 318"
FT 802..813
FT /note="peptide 299"
FT 813..820
FT /note="peptide 268"
FT 836..847
FT /note="peptide R26 2nd sequence"
FT 1010..1031
FT /note="THE TNT repeat 1"
FT 1072..1093
FT /note="THE TNT repeat 2"
FT 1101..1126
FT /note="THE TNT repeat 3"
FT 1158..1183
FT /note="THE TNT repeat 4"
FT 1286..1311
FT /note="THE TNT repeat 5"
FT 1314..1339
FT /note="THE TNT repeat 6"
FT 1349..1374
FT /note="THE TNT repeat 7"
FT 1414..1439
FT /note="THE TNT repeat 8"
FT 1774..1781
FT /note="peptide 293 2nd sequence"
FT 1808..1819
FT /note="peptide 115"
FT 1819..1840
FT /note="peptide 261 1st sequence"
FT 1853..1863
FT /note="peptide 240"
FT 1901..1919
FT /note="peptide R32"
FT 1919..1930
FT /note="peptide 261 2nd sequence"

XX W09413315-A.

XX 23-JUN-1994.

XX 03-DEC-1993; 93WO-US11721.

XX 04-DEC-1992; 92US-0989842.

XX 12-APR-1993; 93US-0046585.

XX (COLD-) COLD SPRING HARBOR LAB.
(TULA-) TULARIK INC.

XX Herr W, Lamarco K, Wilson A;

XX WPI; 1994-234207/28.

XX N-PSDB; AAO69229.

XX New Host Cell Factor polypeptide(s) and nucleic acid - are used
PT to develop agents for diagnosis or treatment of disease
PT associated with expression of a HCF-modulated gene e.g. viral
PT infections

XX Disclosure; Page 39; 71pp; English.

XX HCF is required for the transcription of a number of

XX viral genes, such as the immediate early herpes simplex virus-1

XX genes. Epitopes of the encoded protein can be used in

XX defining functional domains of HCF, identifying compounds that

XX associate with HCF or designing compounds capable of modifying HCF

XX transcription. Such agents can be used to treat viral infections.

XX Sequence 2035 AA;

		9.1%:	Score	204;	DB	15;	Length	2035;
			Best Local Similarity	23.5%:	Pred.	No. 5.1e-06;		
	Matches	119;	Conservative	66;	Mismatches	188;	Indels	134; Gaps
OY	10 VSAPPKVSSGPRLEPAPQAVAKAPTNTTIOFPANILQLPGTVLTKNSNGPLMLVSPOQT	69	:	: ::	::	:: :	::	:
Dd	537 IGSPQSGMALLAAATAAQIKIPPSA--PIIVSVPAAGTIIVKT-----NAVIPTTT	588	:	: ::	::	:: :	::	:
OY	70 TRAFETSNITSRPAVPANPDVTVAICHPNASSOLLIKKNAVTPVKLIAIGITVV----	122	:	: ::	::	:: :	::	:
Dd	589 LPA--TWKVASPPWMSNPAT-----RMLKTA-----AOGVTSVASINTS	628	:	: ::	::	:: :	::	:
OY	123 ---VFYPKKSSVOVNVPTSVMTPRGKRINLVTLK-PSSL-GASSPNSEPM----	173	:	: ::	::	:: :	::	:
Dd	629 TRPIITIHKSGTV-AQAQAVTTYVGCGYTKITIYLKPISVPCGSALISMGLKMYSV	687	:	: ::	::	:: :	::	:
OY	174 --KAENSAVAINLSPTMLENKCKNFL--AMLKLACSGSQSP-----ENGONVK	221	:	: ::	::	:: :	::	:
Dd	688 QTRFVGQSAYTGASSTGPVNIQIQTFGRPLPACTILELVLTSADGKRTIIITTGAAGCTK	747	:	: ::	::	:: :	::	:
OY	222 KLVSQLDANKIEAEFTFRKIYVELKSSPOPHLYPFLLKSVALRKOLLPNSSGFIDCCVQQ	281	:	: ::	::	:: :	::	:
Dd	748 PTILGI-----SSVSPSTP---KPTTIITIKTIMSATITOGATG	785	:	: ::	::	:: :	::	:
OY	282 TSDMVATICTTYVT-----SPVYT-----TYTSSSOSESITVSGA-----TA	321	:	: ::	::	:: :	::	:
Dd	786 VTSSFPIKSPITTIITIKVMTCGTGAPAKKITIAVPAKATIATHGOQGVTQYVLKCAPGPC	845	:	: ::	::	:: :	::	:
OY	322 PRT-----VSVDLTNLPLAGPVGAR--AGVVTLSHSGPATGGTTAGTGLQTS	368	:	: ::	::	:: :	::	:
Dd	846 LRTPMGVRLPYRPVYSAYKPAVTLTVLKGTGVTTLGTGTGYIST--SLAGAGHSITS	903	:	: ::	::	:: :	::	:
OY	369 KPPLYTSVA--NYITYVSLQEKKRVYGSTAVTLSLPADVTEGETSGAICLPSVKRYVSFCW	426	:	: ::	::	:: :	::	:
Dd	904 ASLATPTTLTGATIALTSQQ---VINPFAIWASAQTILLTAGGTLTPITTMQPV-----	954	:	: ::	::	:: :	::	:
OY	427 DHICKPVIGTPOAIKU-NORGPLVISOP	452	:	: ::	::	:: :	::	:
Dd	955 -----SOPTOVTLITAPSGVEARDP	973	:	: ::	::	:: :	::	:
	RESULT 14							
	ABB69806							
ID	ABB69806 standard; Protein; 1795 AA.							
XX	ABBB69806;							
AC								
XX	26-MAR-2002 (first entry)							
DT	Drosophila melanogaster polypeptide SEQ ID NO 36210.							
DE	Drosophila melanogaster developmental biology; cell signalling; insecticide;							
KW	pharmaceutical.							
XW	Drosophila melanogaster.							
SX	wc0200171042-A2.							
FN								
XX	27-SEP-2001.							
PD								
XX	23-MAR-2001; 2001MO-USO9231.							
PE								
XX	23-MAR-2000; 2000US-191637P.							
PR	11-JUL-2000; 2000US-0614150.							
XX	(PEKE) PE CORP NY.							
PA	Venter JC, Adams M, Li PWD, Myers EW;							
FI								
XX	WPI; 2001-656860/75.							
DR	N-PSDB; ABLI3909.							
TX	New isolated nucleic acid detection reagent for detecting 1000 or more							

[illegible]

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:54:50 ; Search time 8.71092 Seconds

(Without alignments)
2152.161 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_452

Perfect score: 2249

Sequence: 1 GTLVTKVAPVSAAPKVVSGP.....VIGTPVOIKLAPGPVLSQP 452

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	2249	100.0	801	T2DT_HUMAN	Q92750 homo sapien
2	521	23.2	1083	T2D3_HUMAN	O00268 homo sapien
3	241.5	10.7	921	T2D3_DROME	P47825 drosophila
4	206.5	9.2	2035	HFCI_HUMAN	P51610 homo sapien
5	206	9.2	2090	HFCI_MESAU	P51611 mesocricetu
6	200.5	8.9	1367	AMVH_YEAST	P08640 saccharomyc
7	195.5	8.7	5179	MUC2_HUMAN	O02817 homo sapien
8	194	8.6	865	CPN2_DROME	O02910 drosophila
9	191	8.5	670	VS0_HSV1	O00130 Ictaelurid h
10	181.5	8.1	662	MUC1_XENLA	O05049 xenopus lae
11	175.5	7.8	3178	YS89_CAEEL	O09624 caenorhabdi
12	172.5	7.7	797	VLGX_HSV8	P28968 equine harp
13	172	7.6	886	VGP3_EBYA8	O07284 epstein-bar
14	172	7.6	1322	YAG3_YEAST	P39712 saccharomyc
15	171.5	7.6	2700	ZAN_HUMAN	O94743 homo sapien
16	170	7.6	1161	DAN4_YEAST	P47179 saccharomyc
17	167.5	7.4	725	AGAI_YEAST	P32323 saccharomyc
18	167	7.4	5376	ZAN_MOUSE	O88799 mus musculu
19	165	7.3	606	SP2_HUMAN	O02086 homo sapien
20	165	7.3	881	YU8_YEAST	P47033 saccharomyc
21	164.5	7.3	1260	ALSI_CANAL	P46590 candida alb
22	164	7.3	1199	N121_RAV	P52591 rattus norv
23	164	7.3	1365	SUZ2_DROME	P51172 drosophila
24	161	7.2	860	CH12_COCIM	P54197 coccoidiolde
25	161	7.2	1119	ALSI_CANAL	O74623 candida alb
26	159.5	7.1	1169	YK82_YEAST	P36170 saccharomyc
27	156.5	7.0	1609	FTG2_YEAST	P25653 saccharomyc
28	156	6.9	1075	FLOS_YEAST	P38694 saccharomyc
29	155.5	6.9	2090	N214_HUMAN	P35658 homo sapien
30	154	6.8	1140	YM96_YEAST	O04965 saccharomyc
31	153	6.8	1306	MSB2_YEAST	P32334 saccharomyc
32	152	6.8	604	MTG8_HUMAN	O06455 homo sapien
33	152	6.8	907	VGP3_EBY	P03200 epstein-bar

34	151.5	6.7	745	1	OCT1_PIG	Q29076 sus scrofa
35	150.5	6.7	743	1	OCT1_HUMAN	P14859 homo sapien
36	150	6.7	1537	1	FLO1_YEAST	P32768 saccharomyc
37	149.5	6.6	676	1	MUC1_MESAU	O60528 mesocricetu
38	149.5	6.6	1036	1	P200_MYCPN	P75211 mycoplasma
39	148	6.6	3866	1	HRX_MOUSE	P55200 mus musculu
40	147	6.5	528	1	PODX_HUMAN	O00592 homo sapien
41	147	6.5	577	1	PODX_MOUSE	O61909 mus musculu
42	147	6.5	1216	1	YPS5_CAEEL	O09277 caenorhabdi
43	146.5	6.5	3969	1	HRX_HUMAN	O03164 homo sapien
44	145.5	6.5	630	1	MUC1_MOUSE	O02496 mus musculu
45	145.5	6.5	2717	1	ZEPI_HUMAN	P15822 homo sapien

ALIGNMENTS

```

RESULT 1
T2DT_HUMAN          STANDARD;          PRT;          801 AA.
ID T2DT_HUMAN
AC Q92750;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription Initiation factor TFIID 105 kDa subunit (TAFII-105)
DE (TAFII105) (Fragment).
GN TAF4B OR TAF2C2 OR TAFII105.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=97011146; PubMed=8658156;
RA Dikstein R., Zhou S., Tjian R.;
RT hTAFII130.";
RL Cell 87:137-146(1996).
CC -1- FUNCTION: CELL TYPE-SPECIFIC SUBUNIT OF TFIID THAT MAY FUNCTION AS
CC A GENE-SELECTIVE COACTIVATOR IN CERTAIN CELLS. TFIID IS A
CC MULTIMERIC PROTEIN COMPLEX THAT PLAYS A CENTRAL ROLE IN MEDIATING
CC PROMOTER RESPONSES TO VARIOUS ACTIVATORS AND REPRESSORS.
CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs). TAFII105 MAY FORM
CC HETERODIMERS WITH TAFII130.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: FOUND IN B CELLS BUT NOT IN OTHER TESTED
CC CELLS.
CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Y09321; CAAT0499.1; -
DR Genew; HGNC:11538; TAF4B.
DR MIM; 601689; -
DR InterPro; IPR003894; TAF_hom.
DR SMART; SM00549; TAF4; 1.
KW SMART: transcription regulation; Nuclear protein.
FT NON_TER
SQ SEQUENCE 801 AA; 85658 MW; D12B493FEA49CD2 CRC64;
Query Match 100.0%; Score 2249; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 7.8e-122;
Matches 452; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 GTLVTKVAPVSAAPKVVSGPRLPAPQIVAVKAPNTTTIQPPANILQLPGTVLINSNSGPL 60
|||||

```

Db 1 GILYTKAVASAPKVSQGRPLPAPQIVANKAPNTTIOEPANQLPFGTVLKNSGRL 60
 QY 61 MLVSPQOYVTRAFETSNITSRPAVPANPQVTKICIVPNSSQLIKKVAATPVKKLAQIGT 120
 Db 61 MLVSPQOYVTRAFETSNITSRPAVPANPQVTKICIVPNSSQLIKKVAATPVKKLAQIGT 120
 QY 121 TVYVTPKPSVQSVAVPTSVYVTPGKPLNTYTTTLKPSLSGSSPSPNPNKAENSA 160
 Db 121 TVYVTPKPSVQSVAVPTSVYVTPGKPLNTYTTTLKPSLSGSSPSPNPNKAENSA 160
 QY 181 VOINLSPTMLENVKCKNFKLMLIKLACSGSQSPENQVKKKVEQLLDKAEFEETRK 240
 Db 181 VOINLSPTMLENVKCKNFKLMLIKLACSGSQSPENQVKKKVEQLLDKAEFEETRK 240
 QY 241 LVEYELKSSQPHLPVPLKKSVAVALROLLENPSQSFIOCCVOQISSDMVIATCTTTVTSPV 300
 Db 241 LVEYELKSSQPHLPVPLKKSVAVALROLLENPSQSFIOCCVOQISSDMVIATCTTTVTSPV 300
 QY 301 VTTTVSSSQSEKSIYSGATAPRTVSQTLNPLAGPVGAKAGVYTLHSVGPATAGTTA 360
 Db 301 VTTTVSSSQSEKSIYSGATAPRTVSQTLNPLAGPVGAKAGVYTLHSVGPATAGTTA 360
 QY 361 GGGLLQTSKPLVTSVANTVTVVSLQPEKPVSGTAVTLSPAVTFGETSGAALCLPSVKP 420
 Db 361 GGGLLQTSKPLVTSVANTVTVVSLQPEKPVSGTAVTLSPAVTFGETSGAALCLPSVKP 420
 QY 421 VVSFCDHICKPVTGTPVQIKLAQPGPVLSQP 452
 Db 421 VVSFCDHICKPVTGTPVQIKLAQPGPVLSQP 452

RESULT 2
 ID T2D3_HUMAN STANDARD: PRT: 1083 AA.
 AC 000268; Q99721; Q9BX42; Q9BR40;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcription Initiation factor TFIID 135 kDa subunit (TAFII-135)
 DE (TAFII135) (TAFII-130) (TAFII130).
 GN TAF4 OR TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI-TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-97336072; PubMed-9192867;
 RA Mengus G., May M., Carre L., Champion P., Davidson I.;
 RT "Human TAFII135 potentiates transcriptional activation by the AF-2s
 of the retinoic acid, vitamin D3, and thyroid hormone receptors in
 mammalian cells.";
 RL Genes Dev. 11:1381-1395(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-21638749; PubMed-11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 Jones M., Scavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
 Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 Graham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,
 Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 Leharajallo M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 Marsh V.L., Martin S.D., McComnachie L.J., McLeay K., McMuray A.A.,
 Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,

RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Seha H.K., Showkneen R., Sims S.,
 RA Skuce C.D., Smith M.T., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tycamore A.C., Vaudin R., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 RN [3]
 RP SEQUENCE OF 105-1083 FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE-97098442; PubMed-8942962;
 RA Tanese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;
 RT "Molecular cloning and analysis of two subunits of the human TFIID
 complex: hTAFII30 and hTAFII100.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).
 CC -1- FUNCTION: MAKES PART OF TFIID IS A MULTIMERIC PROTEIN COMPLEX THAT
 CC PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS
 CC ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION
 CC BY THE AF-25 OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.
 CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
 CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y11354; CAA72189.1; -;
 DR EMBL: AL137077; GAC36006.1; -;
 DR EMBL: AL109911; CAC22312.2; -;
 DR EMBL: U75308; AAC50901.1; -;
 DR TRANSFAC: T02328; -;
 DR GeneW: HGNC:11537; TAF4.
 DR MIM: 601796; -;
 DR InterPro: IPR003894; TAF_hom.
 DR SMART: SM00549; TAFH.1.
 DR KMW Transcription regulation; Nuclear protein.
 FT DOMAIN 39 42
 FT POLY-HIS.
 FT DOMAIN 52 57
 FT POLY-ALA.
 FT DOMAIN 98 101
 FT POLY-GLY.
 FT DOMAIN 142 148
 FT POLY-ALA.
 FT DOMAIN 266 275
 FT POLY-PRO.
 FT DOMAIN 331 337
 FT POLY-ALA.
 FT DOMAIN 680 683
 FT POLY-PRO.
 FT DOMAIN 808 813
 FT POLY-ALA.
 FT DOMAIN 828 831
 FT POLY-ASP.
 FT CONFLICT 105 117
 FT PGSPSRRLPVLA -> GGGLLQQRGRRES
 FT (IN REF. 3)
 FT CONFLICT 136 136
 FT A -> S (IN REF. 2).
 FT CONFLICT 185 185
 FT G -> GPG (IN REF. 2).
 FT CONFLICT 264 264
 FT MISSING (IN REF. 3).
 FT CONFLICT 293 293
 FT P -> L (IN REF. 3).
 SQ SEQUENCE 1083 AA; 109943 MW; A6453827572A0752 CRC64;

Query Match 23.2%; Score 521; DB 1; Length 1083;
 Best Local Similarity 34.0%; Pred. No. 1,1e-22;
 Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

QY 13 PVRVSSG-----PLPAPQIVAVKAPNTTIOEPANQLPFGTVLKNSGRL 61
 Db 414 PRTTSGIRATLTPTVLAPELPQP-----PQNPNTQ--NFQDGVLVSENGQL 464
 QY 62 LVSPQOYVTRAFETSNITSRPAVPANPQVTKICIVPNSSQLIKKVAATPVKKLA 115
 Db 465 MI-PQALAMQMAHNAQPTTMAPRAPRTSAPVQISIVAGPPIIAR-QVTP----- 518

QY 188 TMLNFK-KCKNPLAMLIACSGSPKMGVKKVLEQLDAKTEAEFTKLYVELK 246
 DB 288 TGGNKKCKRCKFLANLIEI--STREPKVEKNVRLIELVANVEPEECRLERL 345
 QY 247 SSFOPHLVPELKKSVYALNQL-----LPNSOSFIO--- 276
 DB 346 ASBPOLCIEFLKKSJLPLNQLATYKELVIEGIRPPONHVLGAGLSQQLPKIQAOIRPIC 405
 QY 277 ---OCVQOTSSDMVATCTTTVTTSPPVTTVSSSQSEKSIIVSGATAPRTVS---VOT 329
 DB 406 PSQTTTIGOTGVAMI--TTPNALGTPTPTIGHTITISKQPPN---IRLPAPRLVNTGIGIRT 460
 QY 330 LNPAGPVCAGKGVYTLHVSPTAAGCTTAGCGLLOTGKPLVTSVANTVYTSLOPEK 389
 DB 461 QLP-SLQVPGQANIVQIR--GPQHAQLQKRGVQVIRATRRP-----PNSVPLN----- 506
 QY 390 VVSGTAVTSLPVPFETGSGAICLPVYK 420
 DB 507 -----KLPAVKVQTOIKAI-TPSLHP 527

RESULT 4
 HFC1_HUMAN STANDARD: PRT: 2035 AA.
 ID HFC1_HUMAN
 AC P51610
 DT 01-OCT-1996 (rel. 34, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE Host cell factor C1 (HCF) (VP16 accessory protein) (HFC1) (VCAF) (CFP).
 GN HFC1 OR HCF1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS), AND PARTIAL SEQUENCE.
 RX MEDLINE=93327419; PubMed=8392914;
 RA Wilson A.C., Lamarco K., Peterson M.G., Herr W.;
 RT "The VP16 accessory protein HCF is a family of polypeptides processed from a large precursor protein.";
 RL Cell 74:115-125(1993).
 RN [2]
 RP SEQUENCE OF 65-2035 FROM N.A. (LONG AND SHORT ISOFORMS).
 RC TISSUE=Retal brain;
 RX MEDLINE=95130107; PubMed=7829097;
 RA Fratini A., Faranda S., Redolfi E., Zucchi I., Villa A., Patrosso M.C., Strina D., Susani L., Vezoni P.;
 RT "Genomic organization of the human VP16 accessory protein, a housekeeping gene (HFC1) mapping to xq28.";
 RL Genomics 23:30-35(1994).
 RN [3]
 RP PROCESSING.
 RX MEDLINE=9603796; PubMed=7590226;
 RA Wilson A.C., Peterson M.G., Herr W.;
 RT "The HCF repeat is an unusual proteolytic cleavage signal.";
 RL Genes Dev. 9:2445-2458(1995).
 CC -1- FUNCTION: UPON LYTIC INFECTION OF PERMISSIVE CELLS, THE HSV TRANSACTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOIF-BINDING PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES.
 CC -1- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM 110 TO 150 KDA AND A MINOR 300 KDA POLYPEPTIDE. THE MAJORITY OF N- AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBEIT NONCOVALENTLY, ASSOCIATED.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A SHORT FORM, MAY BE PRODUCED BY ALTERNATIVE SPLICING. IN THE SHORTER FORM THE N- AND THE C-TERMINAL FRAGMENTS FAIL TO ASSOCIATE.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN FETAL TISSUES AND THE ADULT KIDNEY. PRESENT IN ALL TISSUES TESTED.

CC -1- DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE SIGNAL.
 CC -1- PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS AT A DEFINED SITE, PCE/THET, WITHIN THE HCF REPEAT.
 CC -1- PTM: GLYCOSYLATED. CONTAINS O-LINKED N-ACETYLGLUCOSAMINE RESIDUUS.
 CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).
 CC EMBL: L20010; -; NOT ANNOTATED. CDS.
 DR EMBL: X79198; CA455790.1; ALT_INIT.
 DR Genew; HGNC:4839; HCF1.
 DR MIM; 300019; -
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR001798; Kelch.
 DR Pfam; PF00041; fn3; 1.
 DR Pfam; PF01344; Kelch; 5.
 DR SMART; SM00060; FN3; 1.
 DR Nucleic protein; Repeat; Alternative splicing; Glycoprotein.
 KW REPEAT 44
 FT REPEAT 93
 FT REPEAT 140
 FT REPEAT 148
 FT REPEAT 217
 FT REPEAT 265
 FT REPEAT 313
 FT REPEAT 331
 FT DOMAIN 1010
 FT REPEAT 1010
 FT REPEAT 1035
 FT REPEAT 1072
 FT REPEAT 1097
 FT REPEAT 1101
 FT REPEAT 1126
 FT REPEAT 1158
 FT REPEAT 1183
 FT REPEAT 1286
 FT REPEAT 1311
 FT REPEAT 1314
 FT REPEAT 1339
 FT REPEAT 1349
 FT REPEAT 1374
 FT REPEAT 1414
 FT REPEAT 1439
 FT VARSPLIC 382
 FT CONFLICT 564
 FT CONFLICT 564
 FT CONFLICT 603
 FT CONFLICT 603
 FT CONFLICT 1164
 FT CONFLICT 1164
 FT CONFLICT 1873
 FT SEQUENCE 2035 AA; 208841 MW; 99207FEB875204C0 CRC64;
 SQ

Query Match 9.2%; Score 206.5; DB 1; length 2035;
 Best Local Similarity 23.3%; Pred. NO. 0.00024;
 Matches 135; Conservative 71; Mismatches 194; Indels 179; Gaps 30;

QY 2 TLVTKAPVSAAPPKSSGRLP-----APQIVAKAPMTT-----IQ 39
 DB 446 TLDPQAP-APPPTTTIQLVLPVPGSSISVPAARQGVPAUKVATGATGTPLVM 503
 QY 40 FPAV-----LQLPPTVL-----KNSGPLMLVSPQOT-----VTRAEFTSNT--TS 80
 DB 504 RPASQAGKAPVTSLPAGRVMPVPTOSAGTIGSSPQMSGMAALAAAAAATQKIPSS 563
 QY 81 RP--AVPAN-----PQVTKICVP-----NSSSOLIKKVAVTPVKRLQ 117
 DB 564 RPLVLSVPACTIVTKMAVTPGTTTLPAIVKAVSSPVMNSPNTMLKTA-----AQ 616
 QY 118 IGTIV-----VTVKPPSSVSAVPTSVYTPGKPLNTVYTLK-PSSL-GASS 165
 DB 617 VGTSSVSAATSTTRPITIVKSGTV-TVAQAQVVTTVGCVKRTITLVKSPISVPGGSA 675
 QY 166 TPSENPRL-----KENSAAVOINISPTMLEVKKCKKNFL--AMLIKACSSQSD--- 214
 DB 676 LISNGLKSVSVOTKPKQPSAVTGAOSTGPTVTOITGKPLPAGTILKLTVSADGKPTTI 735
 QY 215 -----EMGQVKKLVLEQLDAKTEAEFTKLYVELKSSQPHLVFLKKSVALNQLP 269


```

DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Calphotin.
GN CPN OR CAP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN (1)
RN SEQUENCE FROM N.A.
RC STRAIN=Canlon-S;
RX MEDLINE=93165729; PubMed=8094559;
RA Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
RT "Calphotin: a Drosophila photoreceptor cell calcium-binding protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).
RN (2)
RN SEQUENCE FROM N.A.
RC STRAIN=Canlon-S;
RX MEDLINE=93165730; PubMed=8434015;
RA Ballinger D.G., Xue N., Harshman K.D.;
RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds
calcium and contains a leucine zipper.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
CC -1- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO
REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL
OF CA+2 PER MOL OF PROTEIN.
CC -1- SUBUNIT: HOMODIMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
CC -1- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
COMPOUND EYES AND OCCELLI.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
DEVELOPMENT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation :
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL: 102111; AAA28405.1; -
DR EMBL: 105080; AAA28420.1; -
DR PIR: A47282; A47282.
DR Flybase: Fbgn0010218; Cpn.
KW Calcium-binding.
FT CONFLICT 36 36 A -> AVAPAVVA (IN REF. 2).
FT CONFLICT 43 43 I -> T (IN REF. 2).
FT CONFLICT 64 64 T -> V (IN REF. 2).
FT CONFLICT 76 76 I -> A (IN REF. 2).
FT CONFLICT 100 100 P -> PP (IN REF. 2).
FT CONFLICT 126 127 VO -> AP (IN REF. 2).
FT CONFLICT 154 154 I -> V (IN REF. 2).
FT CONFLICT 160 160 S -> T (IN REF. 2).
FT CONFLICT 534 534 A -> E (IN REF. 2).
FT CONFLICT 699 699 I -> T (IN REF. 2).
FT CONFLICT 703 703 V -> L (IN REF. 2).
FT CONFLICT 721 721 D -> E (IN REF. 2).
SQ SEQUENCE 865 AA; 84781 MW; 2110417E0B0E7CEE CRC64;

Query Match 8.6%; Score 194; DB 1; Length 865;
Best Local Similarity 22.2%; Pred. No. 0.00045;
Matches 121; Conservative 71; Mismatches 209; Indels 144; Gaps 23;
OY 1 GTLVTKV-APVSAP-----PKVSSGRLPAP-----QIYAVKAPNTTIQFPANIQ 45
DB 4 GTIPSPVASAPVAPVPSAVAPVQVSPAAVAPAPAPAPVAVPVPPLASVQPPVTV 63
OY 46 LPBGVILKINSGLMLVSPQQTVTRAEFTTSNITSRAVPANPQVTKICVPS-SSQLI 104
DB 64 IPAPAPIAAASVTPVASVAP--PVVAAPTP-----PAASPVSTPVAVAQIPVASAPVA 115

```

```

OY 105 KKVAVPVPKVL-AQIGTVTVTPKPSVQSVAVPVSVM-VTPGKPLNTVTLKPSSLG 162
DB 116 PPVAAPPTPVQIPVAPVAVPVIATPPVVASAPTPPAVTPVISPVIASPPVAPANTVPVAP 175
OY 163 ASSTPSNEPMILKAKNSAAVQINLSPTMLE-----NVKCKNPLAMLIKACS 209
DB 176 VAAVPAVPVAVPVLAAVAPVAPVAAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAV 227
OY 210 GSQSPKMGQVKKVLEQLDAKIEAEETRKLYVELKSPDQPLVPLKKSVAVALROL 269
DB 228 -----PEVSVATKRPILAAPVAVVAPATETPPVAPAAASHVASPAVETAVVA-----P 278
OY 270 NSQSFIOCCVOQTSSSDKVIATCTTTTTSPTVTTVSSQSEKSIIV-----SGATAPRV 325
DB 279 VSAS-----TEPPVAAATLTTPAPT-PALAPVVAESQVAAVNTVATPTPAPBETI 329
OY 326 S-----VQTLNPLAGPVGAKGVVTLHSVGTATGTAGTGLQTSKP 370
DB 330 APPVVAETPEVASVAVAEITPPVVPVVAES-----IPAPVAT-----TPVP 372
OY 371 LVTSVANTVTTVTSLOPEKP-----VSGTAVTLT-----LPATTEGET- 408
DB 373 ATLAVTDPDTASAVPELPVIAAPVSAVPAETPVDLAPVLPVVAEVPVAVAEETP 432
OY 409 -----SGAALCPSPKPVVPSFCWD-----HICKVYITGP-----VQIKLAQCP 447
DB 433 ETPAPASAPVTIALDIEVAPVIAAPSDAPAEAPSAADIVSTPTTASVPETTPAPAA 492
OY 448 VLSQP 452
DB 493 VPTEP 497

RESULT 9
VG50_HSV11 STANDARD; PRT; 670 AA.
ID VG50_HSV11
AC 000130;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical gene 50 protein.
GN 50.
OS Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Ictalurid Herpes-like viruses.
OC NCBI_Taxid=10401;
RN (1)
RN SEQUENCE FROM N.A.
RC STRAIN=Aburn 1;
RX MEDLINE=92087490; PubMed=1727613;
RA Davison A.J.;
RT "Channel catfish virus: a new type of herpesvirus.";
RL Virology 186:9-14(1992).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation :
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL: M75136; AAA88153.1; -
DR PIR: F36791; F36791.
KW Hypothetical protein; Repeat.
FT REPEAT 143 158 1.
FT REPEAT 171 186 2.
FT REPEAT 200 214 3.
FT REPEAT 215 233 4.
FT REPEAT 234 252 5.
FT REPEAT 253 268 6.
FT REPEAT 279 293 7.

```


FT REPEAT 294 309 8.
 FT REPEAT 320 334 9.
 FT REPEAT 335 349 10.
 FT REPEAT 362 376 11.
 FT REPEAT 377 391 12.
 FT REPEAT 392 406 13.
 FT REPEAT 407 421 14.
 FT REPEAT 422 436 15.
 FT REPEAT 437 452 16.
 FT REPEAT 464 477 17.
 FT REPEAT 478 493 18.
 FT REPEAT 504 517 19.
 FT REPEAT 518 531 20.
 FT REPEAT 532 545 21.
 FT REPEAT 546 559 22.
 FT REPEAT 560 573 23.
 FT REPEAT 574 587 24.
 FT REPEAT 588 601 25.
 FT REPEAT 602 615 26.
 FT REPEAT 616 629 27.
 SQ SEQUENCE 670 AA; 64174 MW; 2B64A781C519E8B4 CRC64;

Query Match 8.5%; Score 191; DB 1; Length 670;
 Best Local Similarity 24.8%; Pred. No. 0.00049;
 Matches 110; Conservative 41; Mismatches 153; Indels 140; Gaps 22;

QY 16 VSSGRLPAPQIVAKAANTTT-----IOPPANLQPLPGIVLILKNSGPIMLVSPQOVT 70
 DB 248 VTTTPAMPAG---ANDTANITATPTGANDTANTMPAGADIVTTTPAM-----PT 297
 QY 71 RAETTSITSPAVPAMPQVYKICTVNSSQLIKKVAATP-----YKLAQIG 119
 DB 298 GANDTANITTT--ATPAGANDTANTMPAGADIV--VTTTPAMPAGANDTANTMPAGST 353
 QY 120 TTVVTVPKRSPVSVAVPT-----SVTVTPGKP---LNTVTTLKPS-SIGASSTPSNEP 171
 DB 354 DTIVTTTP-----AMPGATDVTTTTPAMPFGATDVTTTPAMPFGATDVTTT 405
 QY 172 NLKANSAAVOINLSPMLLENVKKKFLMLILKACSGSPMCONVKKLVEQLLDK 231
 DB 406 PAKPAGANGVTTTTPAM-----PAGAN----- 428
 QY 232 IEAEFTRLKVELKSSPQHLVFLKSVVALKQLLPNSOSFIQOCVOOT-SSDMYAT 290
 DB 429 -----DVTVTATATPAGANDTANTMPAGADIVTTT 461
 QY 291 CTT--TWTSPVTTTVSSSQSEKSIIVSGAT-APRTVSQVTLNPLAGPVGAKAGVTLH 347
 DB 462 ATVPATGATGVTITTTAKPTGANDTANTMPAGATGVTITTT--AKPTGA-TGIVTVA 516
 QY 348 SVGPATATGTTA-----GTGLQTSKPLVTSVANTVTVLSQPEKPVSGTAVTL 398
 DB 517 TAKPTGATGVTITTTAKPTGANGVTTTAKP--TGATGVTITTTAKPTG--ANGTIVT- 571
 QY 399 SLPAVTFGEISGA--AICLPSVKP 420
 DB 572 -----TTAKPAGANGVTTTAKP 590

RESULT 10
 MUC1_XENLA STANDARD; PRT: 662 AA.
 ID MUC1_XENLA
 AC 005049;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Integumentary mucin C.1 (FIM-C.1) (Fragment).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4; 5; 6 AND 7).
 RC TISSUE-SKIN;
 RX MEDLINE=93077556; PubMed=1447205;
 RA Hauser F., Hoffmann W.;
 RT "P-domains as shuffled cysteine-rich modules in integumentary mucin
 C.1 (FIM-C.1) from Xenopus laevis. Polydispersity and genetic
 RT polymorphism." 267:24620-24624(1992).
 RL J. Biol. Chem. 267:24620-24624(1992).
 CC -1- FUNCTION: COULD BE INVOLVED IN DEFENSE AGAINST MICROBIAL
 CC INFECTIONS. PROTECTS THE EPITHELIA FROM EXTERNAL ENVIRONMENT.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- ALTERNATIVE PRODUCTS: At least 7 isoforms: 1 (shown here), 2, 3,
 CC 4, 5, 6 and 7; may be produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: SKIN.
 CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED.
 CC -1- SIMILARITY: CONTAINS 6 P-TYPE (TREFOLL) DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L02115; AAA74725.1; -.
 DR PIR: A45155; A45155.
 DR HSSP: P01359; 2PSP.
 DR InterPro: IPR000519; P_trefoll.
 DR Pfam: PF00088; trefoll; 6.
 DR SMART: SM00018; P; 6.
 DR PROSITE: PS00025; P_TREFOLL; 6.
 KW Repeat; Amphibian skin; Glycoprotein; Alternative splicing.
 FT NON_TER 1 1
 FT DOMAIN 81 144
 FT 81 144
 FT REPEAT 81 88
 FT REPEAT 89 96
 FT REPEAT 97 104
 FT REPEAT 105 112
 FT REPEAT 113 120
 FT REPEAT 121 128
 FT REPEAT 129 136
 FT REPEAT 137 144
 FT DOMAIN 161 202
 FT 161 202
 FT REPEAT 218 224
 FT REPEAT 225 239
 FT REPEAT 240 249
 FT REPEAT 250 259
 FT REPEAT 260 275
 FT REPEAT 276 287
 FT REPEAT 288 294
 FT REPEAT 295 301
 FT DOMAIN 306 347
 FT 306 347
 FT DOMAIN 353 394
 FT 353 394
 FT DOMAIN 402 522
 FT 402 522
 FT REPEAT 402 411
 FT REPEAT 412 419
 FT REPEAT 420 431
 FT REPEAT 432 443
 FT REPEAT 444 453
 FT REPEAT 454 460
 FT REPEAT 461 472
 FT REPEAT 473 479
 FT REPEAT 480 491
 FT REPEAT 492 498
 FT REPEAT 499 515
 FT REPEAT 516 522
 FT DOMAIN 525 566
 FT 525 566
 FT DOMAIN 572 613
 FT 572 613
 FT DOMAIN 620 661
 FT 620 661
 8 x 8 AA APPROXIMATE TANDEM REPEATS,
 ALA/THR-RICH.
 1-1.
 1-2.
 1-3.
 1-4.
 1-5.
 1-6.
 1-7.
 1-8.
 P-TYPE 1.
 8 x APPROXIMATE TANDEM REPEATS, THR-RICH.
 2-1.
 2-2.
 2-3.
 2-4.
 2-5.
 2-6.
 2-7.
 2-8.
 P-TYPE 2.
 12 x APPROXIMATE TANDEM REPEATS,
 THR-RICH.
 3-1.
 3-2.
 3-3.
 3-4.
 3-5.
 3-6.
 3-7.
 3-8.
 3-9.
 3-10.
 3-11.
 3-12.
 P-TYPE 4.
 P-TYPE 5.
 P-TYPE 6.

```

FT DISULFID 162 188 BY SIMILARITY.
FT DISULFID 172 187 BY SIMILARITY.
FT DISULFID 182 199 BY SIMILARITY.
FT DISULFID 307 333 BY SIMILARITY.
FT DISULFID 317 332 BY SIMILARITY.
FT DISULFID 327 344 BY SIMILARITY.
FT DISULFID 354 380 BY SIMILARITY.
FT DISULFID 364 379 BY SIMILARITY.
FT DISULFID 374 391 BY SIMILARITY.
FT DISULFID 526 552 BY SIMILARITY.
FT DISULFID 536 551 BY SIMILARITY.
FT DISULFID 546 563 BY SIMILARITY.
FT DISULFID 573 598 BY SIMILARITY.
FT DISULFID 583 598 BY SIMILARITY.
FT DISULFID 593 610 BY SIMILARITY.
FT DISULFID 621 647 BY SIMILARITY.
FT DISULFID 631 646 BY SIMILARITY.
FT DISULFID 641 658 BY SIMILARITY.
FT VARSPLIC 240 259 MISSING (IN ISOFORMS 5 AND 6).
FT VARSPLIC 259 259 MISSING (IN ISOFORMS 4 AND 7).
FT VARSPLIC 276 294 MISSING (IN ISOFORMS 4 AND 6).
FT VARSPLIC 278 298 MISSING (IN ISOFORM 5).
FT VARSPLIC 306 350 MISSING (IN ISOFORMS 2 AND 5).
FT VARSPLIC 420 498 MISSING (IN ISOFORM 3).
FT VARIANT 276 276 K -> E.
FT VARIANT 354 354 C -> R.
FT VARIANT 415 415 T -> A.
SQ SEQUENCE 662 AA; 67774 MW; F085227F1ED2FD40 CRC64;

```

Query Match 8.1%; Score 181.5; DB 1; Length 662;
 Best Local Similarity 20.8%; Pred. No. 0.0017;
 Matches 100; Conservative 56; Mismatches 213; Indels 111; Gaps 16;

```

QY 5 TVAP-VPAPVSSGRLPAPQIVAKPNTTIOFPANLQIP-----PGTVLIK 54
DB 70 TTAATATAGAPATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 129
QY 55 SMS-----GPLMIVSPQVTRAEITSNITSRPAVAPNQTAKICVPSNSSLQIK--- 106
DB 130 THSTAAAAPTAASAKSKER--STSSSEEEHCHVPSKREMGCSGKITKCKCKKNC 187
QY 107 -----VAIVPKKLAQIGTIVTTPKPSVQSVAVPTS 140
DB 188 CDPKGGHGHCHHKKHSHHEHTTTKAPTIIQIHTT--TTPPTTTTTPKAPT 245
QY 141 V-----VTVPGRPLNTVTLKPSLSGASTPSNEPIKAE-----NSAAV 181
DB 246 TTTTATPTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 305
QY 182 QINLSTMLENV-----KKCKNPLAMLIKACSGSQSPENQONVKLVEQLLAKTER 234
DB 306 ECKMEPSKREDGYSITESQCR-----KGCDFPSIPQTKWCFYTL--SQVADCKVBP 358
QY 235 EEFTRKLYELKSSPOPHLVPLKKSVA--LRQ---LLPNSOSFIQOQVQOQSSMNTA 289
DB 359 SQ-----RVDCGFKGITADOCROKNCDDSSISGKWKCFYTS--QVAA 400
QY 290 TCTTATVTPSVVTVSSQSEKSIIVSGATAPRVSVOTLPLAGPVAKAGVTLHVS 349
DB 401 TTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTT 453
QY 350 GPTAATGCTAGTGLQTSKPLVTSVANTVTVSLOPEKPVVSGAVTLSLPAVTEGMS 409
DB 454 TPTTTTATATTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTT 513

```

RESULT 11
 VS89_CAEEL STANDARD; PRT; 3178 AA.
 ID VS89_CAEEL 009625; 0969D4;
 AC Q09624; 009625; 0969D4;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)

```

DE Hypothetical protein ZK945.9 in chromosome II.
GN ZK945.9/ZK945.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBL_Taxid-6239;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RC Wilkerson-Sproat J.;
RL Submitted (Feb-1995) to the EMBL/GenBank/DBJ databases.
RN REVISIONS.
RA Durbain R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUPCELLULAR LOCATION: Integral membrane protein (potential).
CC -1- SIMILARITY: CONTAINS 1 GPS DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.cn/announce/
CC or send an email to license@isb-sdb.cn).
CC -----
DR EMBL; Z48544; CAB70192.1; -.
DR EMBL; Z48582; CAB70192.1; JOINED.
DR EMBL; Z48544; CAB70201.1; JOINED.
DR WormPep; ZK945.9; CE25697.
DR InterPro; IPR002111; Cat_channel_TrpL.
DR InterPro; IPR001024; Lipoxigenase_LH2.
DR InterPro; IPR000636; M-channel_nlg.
DR InterPro; IPR000203; PKD_cys-rich.
DR Pfam; PF00520; Ion_transf_1.
DR Pfam; PF01477; PLAT; 1.
DR Pfam; PF01825; GPS; 1.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00308; LH2; 1.
KW Hypothetical protein; Transmembrane.
FT DOMAIN 266 1196
FT DOMAIN 1105 1241 SRR/THR-RICH.
FT DOMAIN 2071 2120 GY/THR-RICH.
FT TRANSMEM 13 30 GPS.
FT TRANSMEM 51 73 POTENTIAL.
FT TRANSMEM 2139 2161 POTENTIAL.
FT TRANSMEM 2348 2367 POTENTIAL.
FT TRANSMEM 2390 2412 POTENTIAL.
FT TRANSMEM 2451 2468 POTENTIAL.
FT TRANSMEM 2483 2505 POTENTIAL.
FT TRANSMEM 2567 2589 POTENTIAL.
FT TRANSMEM 2836 2858 POTENTIAL.
FT TRANSMEM 2939 2961 POTENTIAL.
FT TRANSMEM 2976 2998 POTENTIAL.
FT TRANSMEM 3038 3060 POTENTIAL.
SQ SEQUENCE 3178 AA; 344726 MW; F8239436D03666CD CRC64;

```

Query Match 7.8%; Score 175.5; DB 1; Length 3178;
 Best Local Similarity 22.7%; Pred. No. 0.025;
 Matches 100; Conservative 62; Mismatches 215; Indels 63; Gaps 16;

```

QY 2 TLVTKVAPVAPAPKSSGRLPAPQIVAKPNTTIOFPANLQIPPGVILKSNGLIM 61
DB 356 TLSTSIPTTTTPTTPTTSLSSLPDNLCSYLDYTTTITTTTLMNSTTTEBPSTITTTTE 415
QY 62 LVSPQQTTRAEITSNITSRPAVPA--NPQYKICVTPNSSLQIKKVAVTVPYKLAQIG 119
DB 416 VISTSVTTTTEPTTTLTSTASTSTYEPSTSVTTSPTS-----PVTSTVTS 465
QY 120 TVVTVTPKPSVQSAVA--PTSVTVTPKPLNTVTLKPSLSGASTPSNEPNLKAENS 178
DB 466 SSSSTTVTPTSTSTSTSPSSSTVTTSTAP--STSTGPS--SSSTPSSTAASSSVSST 521

```

```

QY 179 A-AVQINLSPMLMEYKCKFLAMLILKLAGSGSGSPEMOANRKLVEQLDD----- 230
DQ 522 ASYQSSTSYQSSSTTTKSEF-----TTSDDGINDF-YFPEKATTTTDDSTVNL 573
QY 231 -----KIEAEFFTKRYE-----LKSSPQHLVPLKKSVALROLPLNS 271
DQ 574 NSGLIGIYQTSIECTSPTSNSNYSTTKDACEFTKSYSMRLGCTVPASAF-----VGPN 629
QY 272 QSFIOQCVOQTSMDYATCATCTTTTTSPPVTTTYSVSSQSEKSLIIVSGATAPRT-VSQQL 330
DQ 630 YTF---RATMTTDDKKRYYYTANYIODEYSTTIESSSTISAVASSSTSPSPSSTLS 666
QY 331 NPLAGPYGAKGVYTLHSGVFPAATGGTTAGTGLQTSKPLVTSVANTVTVSLQDEKPY 390
DQ 687 STYEPSPSTRSDSTTSAGSTTLQESTT-TSEESTDSDSTTTLIDSTT-SSSPST 744
QY 391 VSGTAATVLSLPATYFGEFTSG 410
DQ 745 ADSTIS-TLSDVDQDFIIDSG 763

```


Db 947 SKSVIPISS--STSGSESESTGASASSSSSSISSEPKSTYSSSLPVTASATSOEI 1004

Oy 397 TSLPAAVTEGTGAACICLPVSKPVSEFCMDHICKPVIGP 438

Db 1005 TSSLPVTTTSTSEOTLV-----TWTSCSESHCTESTISSAI 1041

RESULT 15

ZAN_HUMAN STANDARD; PRT; 2700 AA.

AC Q9Y493; 000218; (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Zonadhesin (Fragment).

GN ZAN.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE OF 1-2379 FROM N.A.

RA MEDLINE=99018118; PubMed=9799793;

RA Glockner G., Scherer S., Schattteoy R., Boright A., Weber J.,

RA Tsui L.C., Rosenthal A.;

RT "Large-scale sequencing of two regions in human chromosome 7q22:

RT analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci

RT reveals 17 genes.";

RL Genome Res. 8:1060-1073(1998).

RN [2]

RP SEQUENCE OF 2338-2700 FROM N.A.

RC TISSUE=Testis;

RA MEDLINE=97271566; PubMed=9126492;

RA Gao Z., Harumi T., Garbers D.L.;

RT "Chromosome localization of the mouse zonadhesin gene and the human

RT zonadhesin gene (ZAN).";

RL Genomics 41:119-122(1997).

CC -1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA

CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR

CC SIGNALING.

CC -1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE

CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).

CC -1- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.

CC -1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE

CC ZONA PELLUCIDA.

CC -1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,

CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF

CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.

CC -1- DOMAIN: THE WMPD DOMAINS 2 AND 3 MAY MEDIATE COVALENT

CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).

CC -1- SIMILARITY: CONTAINS 3.5 MAM DOMAINS.

CC -1- SIMILARITY: CONTAINS 4.5 WMPD DOMAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AF053356; AAC78790.1; -

DR EMBL; U83191; AAC51208.1; -

DR Genew; HGNC:12857; ZAN.

DR MIM; 602372; -

DR InterPro; IPR000561; EGF-like.

DR InterPro; IPR000998; MAM_domain.

DR InterPro; IPR002919; TIL_Cysrich.

DR InterPro; IPR003328; TILA_Cysrich.

DR InterPro; IPR001846; WMP_D.

DR Pfam; PF00094; wvd; 4.

DR Pfam; PF00629; MAM; 4.

DR Pfam; PF01826; TIL; 5.

DR Pfam; PF02345; TILA; 4.

DR SMART; SM00216; WVD; 1.

DR PROSITE; PS01186; EGF_2; 3.

DR PROSITE; PS00740; MAM_1; 1.

DR PROSITE; PS50060; MAM_2; 4.

DR GlycoProtein; Transmembrane; Cell adhesion; Repeat.

FW Glycoprotein; Transmembrane; Cell adhesion; Repeat.

FT NON_TER 1 1

FT DOMAIN <1 109

FT DOMAIN 112 136

FT DOMAIN 161 326

FT DOMAIN 322 446

FT DOMAIN 483 951

FT DOMAIN 953 1065

FT DOMAIN 1066 1454

FT DOMAIN 1455 1861

FT DOMAIN 1862 2292

FT DOMAIN 2293 2684

FT DOMAIN ? ?

FT CARBOHYD 74 74

FT CARBOHYD 403 403

FT CARBOHYD 1023 1023

FT CARBOHYD 1099 1099

FT CARBOHYD 1618 1618

FT CARBOHYD 1737 1737

FT CARBOHYD 1832 1832

FT CARBOHYD 1878 1878

FT CARBOHYD 2136 2136

FT CARBOHYD 2505 2505

FT CARBOHYD 2374 2379

FT NON_TER 2700 2700

SO SEQUENCE 2700 AA; 293013 MW; 80E60CC0B12277B1 CRC64;

Query Match 7.6%; Score 171.5; DB 1; Length 2700;

Best Local Similarity 21.0%; Pred. No. 0.034;

Matches 104; Conservative 60; Mismatches 178; Indels 153; Gaps 22;

Oy 9 PVSAPKVSQPRPA--POLVAVKAPNTTIOFANQLPGRVYLKNSGPIMLVSPQ 66

Db 518 PSEKPMSEKPTIPSEKPTITLTKP-----TIPSEKPTIPSEKPTISTEKPTVPEEP 571

Oy 67 QTVRAETT-----SNITSRPVAPNPQTVK-----IC----- 94

Db 572 TTPEETTTMEPEVPIPEKPSIPEKPSIPEKPTIMEETISTEKPTICPEKPTIPT 631

Oy 95 ---TPNSSQILKAVNPVKK--LAQIGTVVT---TVP--KPS-SVQSVAPPTSVT 143

Db 632 EKPIPEKSTIPSEKPTIPTEKPTIPTEKPTIPTEKPTIPTEKPTIPTEKPT 691

Oy 144 VTGPKPLNTVTLKPS-SLGASSTPSNENPKAEN-SAAVQINSPTMLENVKCKKNFLA 201

Db 692 IPTEKP-TIPTEKPTISTEEPTTPEETTISTEKPSIPMEKPTLP 736

Oy 202 MLIKIACSGSPENQWKKLVEOLDKIAEEFTKLVLELKS---SPQHLVPLK 258

Db 737 -----BETTSVETIISTEKKLTIPEKPTISTEKPTIPTEK 773

Oy 259 KSVVALROLPNOSQFTIOCCVQGTSSDNVIAFTCTVVTSSVVTTVSSSQ---SEKSI 314

Db 774 PTISEKPTIPTEKPT-----LTIPTEKPTIPTEKPTISTEKKLTIPTTEKPT 817

Oy 315 IVSGATAPRTVSVOYLNLAPVGAAGVTLHSVGPRAAGGTAGGLL--QTSKRLV 372

Db 818 I-----SEKPTISTEKPT-----TIPTEKPTIPTEETIISTEKKLTIPEKPTI 860

Oy 373 TSVANTVTVTSIAOPEKPVSGTAVTSLPAAVTEGTGAACICLPVSKPVSEFCMDHICKP 432

Db 861 SPEKLTIP-----EKPTISTEKPTIPTEKLT-----IPTEKPTIP-----TEK 900

Oy 433 VIGTPVOIKLQAPG 447

Db 901 TIPEKLTALRPPH 915

Thu Feb 20 16:36:28 2003

us-09-763-909-2_copy_1_452.rsp

Page 14

Search completed: February 16, 2003, 21:56:12
Job time : 19.7109 secs

C:Genetics:
A:Gene: MUC5B

Query Match 8.2%; Score 185.5; DB 2; Length 3570;
Best Local Similarity 22.0%; Pred. No. 0.0081;
Matches 100; Conservative 53; Mismatches 204; Indels 97; Gaps 14;

QY 2 TLTVKAVAPVAPPPKSSGRLPAPOLVAVKAPNTTIOEPANLQLPCTGLKSNGLPLM 61
DB 3179 TLVLTATATTTGATGAVAPSSSTPGTAHTTKVPTTTTGTATPSSPGTAL-----TPPV 3234
QY 62 LVSPOQTVTRATTSNITSRP-AVPRANPQTVKICVTPNSSSQLIKKVAATPVKKLAQIGT 120
DB 3235 ISFTTPTPTPTPTSGSTVTPSSIPQTHPT-----ARVLT 3269
QY 121 TVTTPVPKSSVQSVAVPTSVVTPGKPLNTVTLKPSLGSSTPSNEPNLKAENSA 180
DB 3270 TTTTTV-----ATGSMATPSSSTQTSCTPPLTTATTATATGCTTPPSSTPGTTPPLPV 3325
QY 181 VOINISPTMLENVKKCNFLAMLIKLACSGSGSPENGQVKKVLEQLDAKIEAEFTTK 240
DB 3326 TSMATTPPATSS-----KANSSSPRTATLPLVLTSTPA--TKSTATST-- 3367
QY 241 LVELKSSQPHLVPELKRKSVVALRQLPNSQSFIOQCVQOTSSDMVATCTT--TVTTS 298
DB 3368 -----PIPSSTLMTWTVP-----QTTPMSTMTSTHTSTPETTHTS 3406
QY 299 PVTTTSSSSOSEKSIIVSGATAPRTVSQVTLNPLAGPVAKAGVYTLHS----- 348
DB 3407 TVLTATATMTFRANSTATPSSTLGTTRILELTTTATTAAATGSTATLSTPCTTWILTE 3466
QY 349 -----VGPAAVGGTTAGTGLQTSKPLVTSVANTVT--TVSLQPEKPVVSGTAVTLSL 400
DB 3467 PSTIATVAVPTGTFATSSSTLGTATHPK--VTAMATMPATATSTVSSSTVGTTTPPAVL 3525
QY 401 PA--VTFGCTSGAICLPSVKPVVSCMDHICKP 432
DB 3526 PSSLPFTSVSTVSSSVLTTLRP--TGFPSSHFTSP 3558

RESULT 15

PC4395
mucin 3 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 10-Nov-1997 #sequence_revision 10-Nov-1997 #text_change 08-Oct-1999
C:Accession: PC4395
R:Van Klinken, B.J.W.; Van Dijk, T.C.; Ounsoren, E.; Buelter, H.A.; Dekker, J.; Elnert
Biochem. Biophys. Res. Commun. 238, 143-148, 1997
A:Title: Molecular cloning of human MUC3 cDNA reveals a novel 59 amino acid tandem repeat
A:Reference number: PC4395; MUID:97445141; PMID:9299468
A:Accession: PC4395
A:Molecule type: mRNA
A:Residues: 1-648 <VAN>
A:Cross-references: DBJ:AF016692; NID:g2454614; PIDN:ABJ71685.1; PID:g2454615
A:Experimental source: Intestine
F:1-59,60-118,119-177,178-236,237-295/Region: repeat
F:296-565/Region: semi-unique #status predicted
F:566-582,583-600,616,617-633,634-647/Region: repeat

Query Match 8.2%; Score 185; DB 2; Length 648;
Best Local Similarity 22.8%; Pred. No. 0.0092;
Matches 102; Conservative 63; Mismatches 188; Indels 94; Gaps 18;

QY 4 VTKVAPVAPVSSGRLPAPOLVAVKAPNTTIOEPANLQLPCTGLKSNGLPLMV 63
DB 137 VDRSTPTVTSQSNRP--TPREVITLPMSTPSEVSTPLT--MPVSTTVTTSSEACTAST 193
QY 64 SPOQTVTRATTSNITSRPVPANPQTVKICVTPNSSSQLIKKVAATPVKKLAQIGTIV 123
DB 194 LVDVDTSTPVTSTQVSSSVTEPEG--TTPIMTISEGS-----IPIT-----IMHVSITRV 242
QY 124 T-----TVPKPSSVQSVAVPTSVVTPGKPLNTVY-----TLKPSLGSSTPSNEPNL 173

DB 243 TSSEGSTLSTPSPVSTPVTTSTEAISSATLSDSTTMSVMPMEISTLGTTLIVSTTPV 302
QY 174 KAENSAVOINISPTMLENVKKCNFLAMLIKLACSGSGSPENGQVKKVLEQLDAKIE 233
DB 303 RPEESTPST--PSV-----YTSMTTASGSSSP-----T 332
QY 234 AEFTTKLVLEKSSQPHLVPELKRKSVVALRQLPNSQSFIOQCVQOTSSDMVATCTT 293
DB 333 TLEGTMTMSTSESTLITVL--ISPISVSPSEASTLSTPPGDDSTPLTSTKAG 389
QY 294 TVTTSFVTTTSSOSEKSIIVSGATAPRTVSQVTLNPLAGPVAKAGVYTLHSVGP 353
DB 390 SFSIPAEVTTIRISITSEK-----TPLTLLVST-----TL-----PTS 424
QY 354 ATGTTAGTGLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAVTSLPVTPESTGA 413
DB 425 FPGASLASTPPLDTSTTPPS-----TDRASTPIVY-----ATTISVYITBESTIGTTI 475
QY 414 CLPSVKPVVSCMD--HICKPVIGTPV 438
DB 476 FIPST--PVTSTADVPFPAATGAVSTPV 501

Search completed: February 16, 2003, 22:00:40
Job time : 25.2925 secs


```

OY 105 KKVAVTTPVKRL-AOIGTVVTVTKRSSVOQSVAVPTSVLT-VTPGKPLMTVTTLKPSISG 162
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 116 PVVAATTPPVQIVPAVPATATPPVAASAPATPAATPVISPIVASPPVPANTTPVPAAP 175
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
OY 163 ASSTPSNEPILKENSAAVOINLSPMLE-----NYKCKNFLAMLKILACS 209
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 176 VAAIPAAVPPVAPVAPVAPVAPVAAVPAETPAPPAEIPVATIPIC-----VAPLI----- 227
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
OY 210 GSQSEAKGQNVKKLIYBDLLDAKIEAEETFRKLYELKSSPOHLVDFLKSSVVALROLLP 269
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 228 -----PEVSVATKPLAAAEPPVVAAPPAETTPVAPAAASPHVSAPVAVETAVAA-----P 278
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
OY 270 NSQSEIOOCVOQOTSMDKVIATCTTTVTTSPVTVTVSSQSEKSIIV-----SGATAPRTV 325
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 279 VSAS-----TEPPVAAATLTTAAET-PALAPVAAEQVAANVAVNTPPIPAPEPETI 329
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
OY 326 S-----VQILNPLAGVYAKAGCVTLHSGVPLAATGGTAGTGLQTSKP 370
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 330 APVPAVPAETPEVASVAAVAAETTPPVVPPVAAES-----IPAPVAT-----TEVP 372
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
OY 371 LVTSVANTVTTVTSLOPEKP-----VSGCTAVTLIS-----LPAPVEGET- 408
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 373 ATLAVTDEBDVTASNVPEPLPVIASPVPSNAVATETPDLARPULPPVAAEPPVAAVAAETP 432
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
OY 409 -----SGAICLPSVKPVVSFCMD-----HICKPVIGTP-----VOIKLAQGP 447
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 433 ETTPAPASAPVTIAALDIPEVAPVIAAPSDAPAPASAAAPIVSTPEPTASVPETAPPA 492
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
OY 448 VLSPQ 452
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 493 VPTPEP 497
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :

```

```

RESULT 10
T34369
hypothetical protein T19D12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34369
R:Favella, A.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid T19D12.
A:Reference number: 221513
A:Accession: T34369
A:Status: preliminary; translated from GR/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-1777 <FAV>
A:Cross-references: EMBL:U041263; PIDN:AAC244.8.1; GSPDB:GN00020; CESP:T19D12.1
A:Experimental source: strain Bristol N2; clone T19D12
C:Genetics:
A:Gene: CESP:T19D12.1
A:Map position: 2
A:Introns: 36/1; 134/2; 180/1; 622/3; 691/2; 754/1; 1111/2; 1174/1; 1271/3; 1322/2; 1681/1

```

Query Match	8.5%	Score 192	DB 2	Length 1777
Best Local Similarity	24.3%	Pred. No. 0.0013		
Matches 106	Conservative	59	Mismatches 176	Indels 96
			Gaps	19

QY	3	LVTKAPVAPSAAPPKSSSGRPLPAPQIVAKAPNTTITQEPANIQLPGGTVLIKNS--GP	59
Db	222	VTTTVAQTTTPAIVTTAN--TTTQGVTTTAGVTTTIVTRAONSTLAAATTPASINTTTCGV	279
QY	60	LMVSPQOQVTRAE-----TTSNITSRPAVPANPQVKTCT-----VPRSS-----S	101
Db	280	TTTVAKTTTVAATQNSTMAATTTANSTPTTQPVVTTSTSTQGSTTAAQTPSSSVLPPTT	339
QY	102	OLIKKVAATVPKKLKLAQIGTIVTTPKRSVQ--SVAVP---TSVATVPEGRKLNTVTLK	157
Db	340	OTTORPTGTGIPSVYTSQGSTSTPISPTTQSSAPSTYSNPTPSTTTLLA--STIA	398
QY	158	PSSLGA--STSPSNEPNLKAENSAVAOINLSPTMELNKKCKNFLMLIKLACSGSOSPEN	216

```

Db      399  PSTGCVPTSSKSSPPN-----SPTPTTTP-----GAPSTL 430
QY      217  GONWKKIVLEQLDAKIEAEETRL-YVELKSSPHLPFLKKSVALRQLLPNOSFI 275
Db      431  GSSSTIVSTTTPS-----TPKASTLTLVSQSTPTSTPLVSS----- 469
QY      276  QOCVOQTSSDMKVIATCTTVTVTTSPPVTVTVSSOSEKSIIVSGATAPRTVSQTLNPLAG 335
Db      470  -----SSSSSSSTVVTSTITPTSTGCVPIST--SNQPTPS--TSNPTTPK--STVTASPTT 519
QY      336  PVGAKAGVYVTHSGPPLAAGGTTAGGLQTSKPLVTSVANTVTYVLSIQEPKPVVSGTA 395
Db      520  GATSTASPTSTTSAPTSPHSPPSS---TMTSTVPVTSFSTASPTTVPI---TPVAGOC 572
QY      396  VTLSPAVTFGÉTSGAA 412
Db      573  YCQSNVAVAPELTSGTS 589

```

RESULT 11
 F36791
 hypothetical protein ORF50 - Ictalurid herpesvirus 1 (strain atburn 1)
 C:Species: ictalurid herpesvirus 1
 A:Note: host Ictalurus punctatus (channel catfish)
 C:Date: 17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change 31-Jan-2000
 C:Accession: F36791
 R:Davidson, A.J.
 submitted to GenBank, January 1992
 A:Description: Channel catfish virus: a new type of herpesvirus.
 A:Reference number: A36804
 A:Accession: F36791
 A:Molecule type: DNA
 A:Residues: 1-670 <DAV>
 A:Cross-references: GB:M75136; NID:g331209; PIDN:AAA8153.1; PID:g331260
 R:Davidson, A.J.
 Virology 186, 9-14, 1992
 A:Title: Channel catfish virus: a new type of herpesvirus.
 A:Reference number: A39447; MOID:92087490; PMID:1727613
 A:Contents: annotation
 A:Note: neither protein nor nucleic acid sequence is given
 C:Genetics:
 A:Gene: 50
 C:superfamily: period clock protein; EGF homology

	Query Match	8.5%	Score 191;	DB 2;	Length 670;
	Best Local Similarity	24.8%;	Pred. No. 0.00042;		
	Matches 110;	Conservative 41;	Mismatches 153;	Indels 140;	Gaps 22;
Oy	16	VSSGRRLPAPQIVANKANTTT-----IQFANLQPPGCTVILKNSGPIMLVSPQOYVT	70		
Db	248	VTTTPPAMPAG---ANDTANITTTATPTGANDRANVTMPAGADTIVTTTPAM-----PT	297		
Oy	71	RAEETSNITSPRAVANDQYKICVPNASSOLIKKAVTP-----YKKLAQIG	119		
Db	298	GANDANITTT--APFAGANDTANVTMPAGADTIV--VTTTPMPAGANDTANVTTPASST	353		
Oy	120	TTVTVTYVKKPSSVOSVAAPT-----SVYVTPCKP---INTVYTLKPS--SLGASSTPSNEP	171		
Db	354	DTIVYTTTP-----AMPTGATDTIVTTTPAMPFGADTIVTTTPAMPFGATDTIVYTTT	405		
Oy	172	NKAEENSAVOINISPTMLLENYKKCKNFLAMLILKLAGSGSPENKGVKKLVEOLDPAK	231		
Db	406	PAKPGANGTIVTTTPAM-----PAGAN-----	428		
Oy	232	IEAEETFRKLYVEIKSSPOPHLVPFLKKSVALRQLLPNSOSFIQOQOQT--SSDMVIAT	290		
Db	429	-----DTVYTTAPATPAGANDTANVTKPTGATDTIVYTT	461		
Oy	291	CTT--TVTTSPTVYTTVSSSSOSEKSIIVSGAT--APRTVSVQTLNDPLAGPVGAKACVTLH	347		
Db	462	ATVKKPTGATGIVTTTAKPTGANDTANVTKPTGATGIVTTT---AAKPTGA--TCTVTVVA	516		
Oy	348	SVGPAPATNGGTTA-----GICGLLOTSKRPVTVSANVTYVSLQDEKPYVSGTAVTL	398		

Query Match	8.9%	Score 200.5;	DB 1;	length 1367;
Best Local Similarity	22.2%;	Pred. No. 0.00028;		
Matches 113; Conservative	69;	Mismatches 225;	Indels 101;	Gaps 17;

```

QY      2 TLTKVAVSAPVSPVSSGRLPLAP-----QIVAKAPNTIOFPALQDPEPTVIK 54
Db      618 TTESSAPVATSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTE 677
QY      55 SNSGPLMLVSPDQT--VTRAEFTSNITSHPAVPANPQVIKICVPPNSSOLIKKVAV- 109
Db      678 SSSAPVATSTTESSAPVATSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPV 737
QY      110 -TPAKKILAOIGTVVVTN-----VPRSSVQS-----AAVPTSVTVTPCKPLNF- 152
Db      738 PTPSSSTTESSAPVATSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTP 797
QY      153 -----VTLTKRSLGASTPEPSNEBPLKAENSAVAQINLSPMLNKKCKNFIA 201
Db      798 SSSSTTESSAPVPTPSSSNITSSAPSSPFPSSSTRESSVP--PTPSSSTESS-- 851
QY      202 MLIKLAGSAGS-----DEMGNKKVLEQLOLAKIAEFTRKIVYELKSSQPHLV 254
Db      852 --APVSSSTTESSAPVPTPSSSNITSSAPSSIPPTSTESFGTIVTPBSSK----- 904
QY      255 PFLKRSVALROLPLNSOSFIOQVOOATSDMVIATCTTIVTSP--VTTTVSSSOS 311
Db      905 -----YPGSQT--ETSVSTETTVTPKTTVTSTPTSTITITTVVCSGTGN 949
QY      312 KSIIVSATAPRIV--SVQJLNLADPVGAKAGVYLIHSGFATNGGTAGTG--LLQT 367
Db      950 SAGETTSGCSCSKVTVTVPTTTTTTVTSSITTTTTTVVCSGTGNSAGETTSGCSPKTTT 1009
QY      368 SKPLVTSVANP---VTVVSLQPEKPVNSGATVATLISLPAVT--FGETSGAIC----- 414
Db      1010 TVPSTSPSEIASBSTTSTPTTVTVVSVTVVTEBYSSTKRGEDITTTTPTKNIPTTY 1066
QY      415 -----LPSYKPVVSRCHMDICKPVYGT 436
Db      1070 LTTIATPSPSVATVNTPTTTTTTVVCSST 1097

```

```

Matches   11%, Conservative    71; Mismatches 206; Indels 139; Gaps 23;

Oy      3 LVTKYAVAPSPKVSQSGRLPAPOIVAKVADNTTTIOFPANLQLPQCTVLIKNSGGLML 62
          ||| : ||||| : 
Db      29 VYSPAAYVAWVAAPAAAPTAVTPVPPTPLASVO- PALTVPAPAFLIAAASVAPAS 87
          ||| : ||||| : 

Oy      63 VSPOQTVRRAETTSNITSRPAVPANPQTKICTVPNS-SSOLLKKVVAVT-----PVKKLA 116
          ||| : ||||| : 
Db      88 VAF--PVYAALFTPP-----AASPSTPPVAAQIPVANSAVAPVATPTPPAPIFVAA 140
          ||| : ||||| : 

Oy     117 QIGITV-VTVTVPKRSSVQSVAVFPTSVTTVPGKPLMTVTTLKSSLGASTGSNEENLK 174
          ||| : ||||| : 
Db     141 PVIAAPPVAASAAPPAAATPVVSP--VIATPPVVPANTTV---PVAAPVAAPVAAPVVA 195
          ||| : ||||| : 

Oy     175 AENSANAQINISPTMLE-----NVKKCKNFLAMLIKILACSGSOSPMEGNVK 221
          ||| : ||||| : 
Db     196 PVLPAAVAPAPAVAPEAEPPAPPPVAELIPVATIEPC--VAPLI-----PEVSVAAT 243
          ||| : ||||| : 

Oy     222 KLVBOLDLDAKIIEAEFEFKKLVELKSSPQHLYPFLLKSYSVALDQLPNDSGFQQCVOQ 281
          ||| : ||||| : 
Db     244 KPLAAEPVVAAPPATETPVVAPAAAASHSVAAFAVEATAVA-----FVASA----- 290
          ||| : ||||| : 

Oy     282 TSDDAVIATCTTGYTVTTSPTVTTTSSSQSEKSIIV---SGATAPRTVS----- 326
          ||| : ||||| : 
Db     291 TEPPAAATLTTLTAPEET-PALAHPVAESQVAAINTYVATPPPTAPEETIIAPVVAETEPEVA 349
          ||| : ||||| : 

Oy     337 ---VOLNLPIAGPYGAKAGVYLHSVGPTAATGGTAGTGLOTLSKPPLYTSVANTTYTV 382
          ||| : ||||| : 
Db     350 SVAAVETPPVPYPVPAABS-----IPAPVAT-----TPPATLATVDPDVTA 392
          ||| : ||||| : 

Oy     383 SLOPEKP-----VSGTAVTLS-----LPADVTEGM-----SC 410
          ||| : ||||| : 
Db     393 SAVPELPVVIAPSPVPSVAETAETPDALAPVLPVPAABEVPVAVVEETPETPAPASAVTI 452
          ||| : ||||| : 

Oy     411 AAICLPSPKVPSFCMD-----HICKRVGTGP-----VOIKLAGPVPVLSOP 452
          ||| : ||||| : 
Db     453 AALDIPEAPVIAAPSDAPAPAPSAAPIYSTPTTASVPETTAPPAAVPTPEP 505
          ||| : ||||| : 

RESULT 9
A47282
calcium-binding protein calphotin - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 21-Sep-1993 #sequence.revision 18-Nov-1994 #text.change 21-Jul-2000
C:Accession: A47282
R:Martin, J.H.; Benzer, S.; Rudnicka, M.; Miller, C.A.
Proc. Natl. Acad. Sci. U.S.A. 90, 1531-1535, 1993
A>Title: Calphotin: a Drosophila photoreceptor cell calcium-binding protein.
A:Reference number: A47282; MUID:93165729; PMID:8094559
A:Accession: A47282
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-865 <MAR>
A:Cross-references: GB:L02111; NID:g157031; PIDN:AAA26405.1; PID:g157032
A:Experimental source: Photoreceptor cells
A>Note: sequence extracted from NCBI backbone (NCBIN:124955, NCBIR:124956)
C:Genetics:
A:Gene: FlyBase:Cpn
A:Cross-references: FlyBase:Fbgn0010218
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
C:Keywords: calcium binding

Query Match           8.6%; Score 194; DB 2; Length 865;
Best Local Similarity 22.2%; Pred. No. 0.00038;
Matches 121; Conservative 71; Mismatches 209; Indels 144; Gaps 23;

Oy      1 GTLVTRYV-APVSAP-----PKVSSGRRLAP-----QIVAANKPNTTTIQPANLQ 45
          ||| : ||||| : 
Db      4 GTISPVSAPVAAPVTSBAVAAPGVQVSPAAVADAPAPAIATVVPAPPTLASAQPRVT 63
          ||| : ||||| : 

Oy     46 LPPEGVLIKNSGGLMLVSPQQTTRAEETSNIISRPAVPANPQTKICTVPNS-SSQLT 104
          ||| : ||||| : 
Db     64 IPAPAPIAASAVTVASVAP--PVYAAPT-----PASFSVSTPVVAQIPVAASAPVA 115
          ||| : ||||| :
```


Db 461 QIP-SIQVPGQNIQVIR--GPOHQLQRTGSVOIRATIRP-----PNSVPTAN----- 506
 Qy 390 VVSGTAVTLSPAVTGETSGAICLPSVKP 420
 Db 507 -----KLTAVKVGQTOIKAI-TPSLHP 527

RESULT 2

A5183
 TBP-associated factor TFIID - fruit fly (*Drosophila* sp.)

C:Species: *Drosophila* sp.

C>Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 16-Feb-1997

C:Accession: A45183

R:Hoey, T.; Weitzierl, R.O.; Gill, G.; Chen, J.L.; Dynlacht, B.D.; Tjian, R.

Cell 72, 247-260, 1993

A:Title: Molecular cloning and functional analysis of *Drosophila* TAF110 reveal properties

A:Reference number: A45183; MUID:93145526; PMID:7678780

A:Accession: A45183

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-921 <HOE>

A:Experimental source: embryo

A>Note: sequence extracted from NCBI backbone (NCBIP:123832)

C:Genetics:

A:Gene: flybase:Taf110

A:Cross-references: Flybase:FBgn0010280

Query Match 10.7%; Score 241.5; DB 2; Length 921;

Best Local Similarity 24.8%; Pred. No. 5.6e-07;

Matches 112; Conservative 65; Mismatches 163; Indels 111; Gaps 17;

Qy 25 POLVAKANNTTITOPANLQLPCTVLIKNSGPM-----VSPQQTVAETTS 76
 Db 133 POSPTLSTLNTGOTPA-----LVKTIDNGFOLLVGTGTPPTVQTITNTSNN 184
 Qy 77 NTSRAVPANPQTVKICVVPNSSQ-----LIKVAVTPVKKLQIGTTVTTP 127
 Db 185 NNTSTTNHTTTO-ILQTVPAASMTNTATSNITVNSVASSGAVNSOPPHLQLNQ 243
 Qy 128 KSSVQSVAVPTSVTVTGKPLNTTLKPSLSGASTPSNEPNLKAENSAVQINLSP 187
 Db 244 AQLPQITQITIPAOOSQOQVNNVSSAGTATAVSTTA-----ATT 287
 Qy 188 TYLENWK-KCKNPLMLIKLAGSGSPKMGONVKKLVQOLDLAEAEFTRKLYVEIK 246
 Db 288 TQGNKKECKREFLANLIL--STREPKVEKNVRLIQLVNAVNEPEFCRLERLIN 345
 Qy 247 SSPPHLVPEFLKSVVALROL-----LPNSQSFQ----- 276
 Db 346 ASQPELIFLKLKSLPLKQALYTKELVETGKPPQHVGLAGLSQQLPKIQALRPIG 405
 Qy 277 ---OCVOQTSSDMVIACTTITVTSPVTTVSSQSEKSIIVSGATAPRTVS---VOT 329
 Db 406 PSQTTTIGOTVYRMI--TINALGTPTPTIGHTTISKQPN---IRLPAPRLVNTGICRT 460
 Qy 330 LNPAGPACAKGVVTLHSGVPLATGTTAGTGLQTSKPLVTSVANTVTYSIQPEK 389
 Db 461 QIP-SIQVPGQNIQVIR--GPOHQLQRTGSVOIRATIRP-----PNSVPTAN----- 506
 Qy 390 VVSGTAVTLSPAVTGETSGAICLPSVKP 420
 Db 507 -----KLTAVKVGQTOIKAI-TPSLHP 527

RESULT 3

T33369

hypothetical protein H02F09.3 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2000

C:Accession: T33369

R:Geisler, C.; Harmon, G.

submitted to the EMBL Data Library, July 1998

A:Description: The sequence of *C. elegans* cosmid H02F09.

A:Reference number: Z21330

A:Accession: T33369

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1275 <GEI>

A:Cross-references: EMBL:AF077538; PIDN:AC64622.1; GSPDB:GN00028; CESP:H02F09.3

A:Experimental source: strain Bristol N2; clone H02F09

C:Genetics:

A:Gene: CESP:H02F09.3

A:Map position: X

A:Introns: 42/1; 78/3; 106/3; 135/2; 161/2; 224/1

C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosida

Query Match 9.6%; Score 215; DB 2; Length 1275;

Best Local Similarity 23.6%; Pred. No. 3.4e-05;

Matches 110; Conservative 71; Mismatches 182; Indels 104; Gaps 18;

Qy 2 TLVTKAPVSAAPKVGSGP---RLPAPQIVAKADNTTITOPANLQLPCTVLIKSN- 56
 Db 275 TVVPTTITVGTPTTVTVPTVTTIPSTVTSITPTSTVTVVPSVTVVPSVATVPKPS 334
 Qy 57 ---SGPLMVS-POQTVTRATTSNITSRAVPANPQTVKICVVPNSSQILKRVATPVK 113
 Db 335 VVAPSTVTVVSTVVTAKPNTV--VTSPTVATPTTV--VTPST-----VTVPS 382
 Qy 114 KLAQIGTTV---TVTPRSSVQSVAVPTSVT-----VTPGKPLNTV 153
 Db 383 TVTVPTVTVTNTSTVTVATPSTV--VTPPTVMTSSTVITPTTGSSPSTAGTSLAST 440
 Qy 154 TLTKPSSLCAGSTSPSNEPNLKAENSAVQINUSPMLNWKCKKNPLMLIKLAGSGOS 213
 Db 441 AVTETSISSSTP-----LPSSQSTLSMSLSSTVPS-----STAGATS 481
 Qy 214 PEKGVNKKLVLEQLDA--KIEAEFTRLVLEKSSQPHLVPEFLKSVVALROLPPS 271
 Db 482 PATQOSTKPTIGSMSSGPTTVAPGASTSTVLAQSTPGTTV-----TLPSG 529
 Qy 272 QSEFIQCVQOOTSMDVIACTTITVTSPVTTVSSQSEKSIIVSGATAPRTVSQTLN 331
 Db 530 SSTPATGTSQASTVTVTVTDISTVSGSVTSQAESSLSTESPTSGSSI-STVSIVSSQ 588
 Qy 332 P-----LAPVCAKAGVTLHSGVPLATGTTAGTGLQTSKPLVTSVA 376
 Db 589 PSTYIPVSSASSISTVLSGSTGTA-----SPGTRESSGSSSTGSPSTISGS-----A 636
 Qy 377 NPTVTSLOPEKPVSGTAVTLSPAVTGETSGAICLPSVKPVYS 423
 Db 637 STVTG-STVTEASTIGSTESTIIPGSTSTVSEASTVSGSSVSTVS 682

RESULT 4

I47141

gastric mucin (clone PGM-2A) - pig (fragment)

C:Species: *Sus scrofa domestica* (domestic pig)

C>Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 03-Nov-2000

C:Accession: I47141; S55315

R:Turner, B.S.; Bhaskar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; Lakont, J.T.

Gastroenterology 106, 200, 1994

A:Title: Pig gastric mucin: isolation and characterization of a cDNA clone with a nov

A:Reference number: I47141; MUID:94102478; PMID:7506218

A:Accession: I47141

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-528 <TUR>

A:Cross-references: EMBL:U10281; NID:9915205; PIDN:AA648526.1; PID:9915208

R:Turner, B.S.; Bhaskar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; Lakont, J.T.

Biochem. J. 308, 89-96, 1995

A:Title: Isolation and characterization of cDNA clones encoding pig gastric mucin.

A:Reference number: S55315; MUID:95275264; PMID:7755593

A:Accession: S55315

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-528 <TUR>

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:55:19 ; Search time 15.2925 Seconds

(without alignments)
2841.441 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_452

Perfect score: 2249

Sequence: 1 GILVTKVAIVSAPKPKVSSCP.....VIGTPOIKLAPGVLSPQ 452

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	241.5	10.7	921	2	A48184
2	241.5	10.7	921	2	A45183
3	215	9.6	1275	2	T33369
4	209.5	9.3	528	2	I47141
5	204	9.1	2035	2	A40718
6	202	9.0	3020	2	A43932
7	200.5	8.9	1367	1	S48478
8	197.5	8.8	873	2	A47283
9	194	8.6	865	2	A47283
10	192	8.5	1777	2	T34369
11	191	8.5	670	2	F36791
12	190.5	8.5	825	2	T29634
13	186.5	8.3	873	2	F96615
14	185.5	8.2	3570	2	T45025
15	185	8.2	648	2	PC4395
16	183.5	8.2	2232	2	T34434
17	182.5	8.1	2187	2	T30826
18	181.5	8.1	662	2	A45183
19	179.5	8.0	867	2	T45463
20	179.5	8.0	1324	2	S52863
21	177.5	7.9	3507	2	T34513
22	175.5	7.8	796	2	T21460
23	174	7.7	549	2	C87719
24	173	7.7	1630	2	A53577
25	172.5	7.7	797	1	VGBEX1
26	172	7.6	886	2	S29605
27	171	7.6	1032	2	T45463
28	171	7.6	1032	2	T45463
29	170	7.6	1161	2	S57180

30	169.5	7.5	1151	2	T18535	high molecular mas
31	169	7.5	1367	2	S51959	hypothetical prote
32	167.5	7.4	725	2	A41258	a-agonist core
33	167	7.4	798	2	T34248	hypothetical prote
34	167	7.4	5376	2	T42215	zonadhesin - mouse
35	166	7.4	851	2	T22696	hypothetical prote
36	165	7.3	881	2	S56032	probable membrane
37	164.5	7.3	752	4	A57784	AMU1/MTG8 mutant f
38	164	7.3	1199	2	A40670	nuclear envelope p
39	164	7.3	1365	2	S14871	suppressor two of
40	163	7.2	896	2	T22061	hypothetical prote
41	162.5	7.2	1260	2	S60896	agglutinin-like pr
42	161.5	7.2	1832	2	T31113	mucin-like glycop
43	161	7.2	2422	2	T12687	ALR protein homolo
44	160.5	7.1	780	2	A48143	HF-1 regulatory el
45	160	7.1	860	2	JC4566	chitinase (EC 3.2.

ALIGNMENTS

RESULT 1

transcription initiation factor IID 110k chain - fruit fly (Drosophila melanogaster)
A:Accession: A48184
C:Species: Drosophila melanogaster
C:Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
R:Kokubo, T.; Gong, D.W.; Roeder, R.G.; Horikoshi, M.; Nakatani, Y.
Proc. Natl. Acad. Sci. U.S.A. 90, 5896-5900, 1993
A:Title: The Drosophila 110-kDa transcription factor TFIID subunit directly interacts
A:Reference number: A48184; MUID:93317591; PMID:8327460
A:Accession: A48184
A:Status: preliminary
A:Molecule type: mRNA; protein
A:Residues: 1-921 <KOK>
A:Cross-references: GB:563550; NID:9398432; PID:9398433
A:Experimental source: embryo nuclear extract
A:Note: Sequence extracted from NCBI backbone (NCBI:134863, NCBI:134864)
C:Genetics:
A:Gene: FlyBase:Tafl10
A:Cross-references: FlyBase:Fbgn0010280
C:Keywords: transcription initiation

Query Match	Best Local Similarity	Score	DB 2;	Length	921;
Matches 112;	Conservative 65;	Mismatches 163;	Indels 111;	Gaps 17;	
QY 25	POIYAVKAPNTTTOFPANLQLPPTVLIKNSGPIML	10.78;	Score 241.5;	DB 2;	Length 921;
DB 133	POSPITLSTINTGOTPA-----LLVKTDNGQLRVGTTGPPVTQTTTNSNS	10.78;	Score 241.5;	DB 2;	Length 921;
QY 77	NITSRAVPANPQVKTCTVENSQ-----LIKKVAVTPVKKLAQIGTVVTVVP	10.78;	Score 241.5;	DB 2;	Length 921;
DB 185	NTSTNTNPTTTO--IRLOTVAASMTTATNSIVASVSGVANSOPHLLQNAQ	10.78;	Score 241.5;	DB 2;	Length 921;
QY 128	KPSVQSVAVPTSVVTVPGKPLNTVTTLAKPSSLCAGSTPSNEPLAENSAVAQINLSP	10.78;	Score 241.5;	DB 2;	Length 921;
DB 244	APOLFOITQIOTIPAQSOQOOVNVSSAGTAAVASTTA-----ATT	10.78;	Score 241.5;	DB 2;	Length 921;
QY 188	TMLENVK-KCKNPLAMLIKACSGSQSPENMGONKVLVEQLLDKLEAEFTKLYELK	10.78;	Score 241.5;	DB 2;	Length 921;
DB 288	TOGQTKCKKFLANLEL--STREPKVKKVNTLQELVANVEEFCDRLERLN	10.78;	Score 241.5;	DB 2;	Length 921;
QY 247	SSPQHLVPLFKKSVALROL-----LPSQSFQ---	10.78;	Score 241.5;	DB 2;	Length 921;
DB 346	ASPPCLIGFLKSLPLRLALYKELVIEIKPPPHVGLAGLSQPLPIQAOIRIG	10.78;	Score 241.5;	DB 2;	Length 921;
QY 277	---OCVOQSSDNVIACTTCTTSPVVTTVSSSQSEKSIIVSGATAPRTVS--VQT	10.78;	Score 241.5;	DB 2;	Length 921;
DB 406	PSQTTTIGOTGVRI--TPNALGTPRPTIGHTTISKPPN--IRLPAPRLVNTGIRT	10.78;	Score 241.5;	DB 2;	Length 921;
QY 330	LNPAGPVGAGAGVTVLHSGVPTAATGTTAGTGLQTSKRLVTVSVANTVTVSLOPEKP	10.78;	Score 241.5;	DB 2;	Length 921;

This Page Blank (uspto)

```

Db 414 PTATSGIRATLPTVLAPRLPOP-----PONPTNIQ---NFOLPGMVLVRSNGQL 464
QY 62 LVSPQOQYTR-----AETTSNITSRAVPANPQVTKICTVPNSSQLKKVAVTPVKKL 115
Db 465 MI-PQOALAQOAHQAHAPQOTMAPRPAFTPSAPVQISTVQAPGPTIAR-QVTP----- 518
QY 116 AQIGTVTVTVKPPSSVQSVAVPTSVTVTPGKPLNTVT--TLKPSLGAAS-----TPS 168
Db 519 -----TTIKOY---SQOQTVQPSKTLQSPGVQVQPLVGGAAQTASLGTANAVQGTIPQ 571
QY 169 NE-PNLKANSAAVOINLSPTMLENVKCKNFAMLIKACSGSQSPMGQNVKKLVEQL 227
Db 572 RTVPQATTTSSAATE-----TMENVKCKNFSLTLIKLASSGKQSTETJANVKELVQNL 625
QY 228 LDKIEAEFEFTRKLYVELKSSPQPHLPFLKKSVALRQLLPNSQSTQOCVQO-----TS 283
Db 626 LDGRIEAEDEFTRSLYRELNSSPOPLVPFLKRSLEPALROLTPDSAAFTQOSQOQPPPTS 685
QY 284 SDWVIATCTTVTVSPVTVTVSSSQSEKSIIVSGATAPRTVSVQTLNPLAGPVGAKAGV 343
Db 686 Q-----ATTALNAVVLSSSVQRTAGKTAATVTSALQPPVLSL----- 722
QY 344 VTLHSVGPAAAGTGTAGTGLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAVTLSPAV 403
Db 723 -----TQPVGVGVGKQGOPTPLVIO-----QPKP----- 747
QY 404 TFGETSGAICLPVKPVPVSCMDHICKPVI--GTPV-----QIKLAOPG 446
Db 748 -----GALIOIOLNLPQPV-----PVVKPAVLPGTKALSAVSAQAAAAQKNKKEPG 792

```

Search completed: February 16, 2003, 22:04:22
 Job time : 20.3897 secs


```
Db 530 LQKIEAEFTSKLYRELNSPOPIVLPFLKSLPALROLTPDSAAFIOOQOQPPPPS 589
Qy 284 SDWVATCTTAVTSPVTTTSSOSEKSIIVSGATAPRTVSQTLNPLAGVGAAGV 343
Db 590 Q-----ATTALTAVLSSVQRTAGKTAATVTSALQPPVLSL----- 626
Qy 344 VILHSVGPAAAGCTTAGGGLQTSKPLVTSVANTVTVSSLOPEKRVSGTAVTSLPAV 403
Db 627 -----TOPQVGVGKQGPPLVIO-----QPPKP----- 651
Qy 404 TGETSGAALCLPSVYVFCMDHICKPVI--GTPV-----QIKLAORG 446
Db 652 -----GALIQLNPLQPV-----PYYKPAVLPGTKALSAVSAQAAAAQKNKLEBG 696

RESULT 13
US-09-724-676-63923
; Sequence 63923, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63923
; LENGTH: 775
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-63923

Query Match
Best Local Similarity 32.6%; Score 507.5; DB 5; Length 775;
Best Local Similarity 32.9%; Pred. No. 1.3e-20;
Matches 157; Conservative 58; Mismatches 121; Indels 141; Gaps 20;

Qy 13 PKVSSG-----PRLAPQIVAVKAPRTTITQPPANIQPPTVILKNSGPI 61
Db 382 PRATTSIGRATLPVLARLP-----PONPTNIQ---NFOLPPGVLVSENGQL 432
Qy 62 LVSPQOVTNR-----AETTSNITSRAVPANDQVTKICTVPSNSQLIKKVAVTPVK 115
Db 433 MI-PQALAQMOQAHAQQTMAAPRATPTSAAPVQISTVOAGTPIIAR-QVTP----- 486
Qy 116 AQIGTVVTVPKPSSVQSAVPTSVVTVPGKPLNTVT--TLKPSLSGASS-----TPS 168
Db 487 -----TTIIQV---SQAQTTVQPSATLQSPGVQPLVGGAAQTASLGTATAVQGTQ 539
Qy 169 NE-PNLKANSAAVQINLSPTMLENKKCKNFLAMLIKILACSGSGSPENQVKKLVEOL 227
Db 540 RTVPGATTTSSATE-----TMENYKCKCKNFLSTLIKILASSGKOSTETAAVKEIYQNL 593
Qy 228 LQAKIAEFTSKLYRELNSPOPIVLPFLKSLPALROLTPDSASFIOOQVQO-----TS 283
Db 594 LQKIEAEFTSKLYRELNSPOPIVLPFLKSLPALROLTPDSAAFIOOQOQPPPPS 653
Qy 284 SDWVATCTTAVTSPVTTTSSOSEKSIIVSGATAPRTVSQTLNPLAGVGAAGV 343
Db 654 Q-----ATTALTAVLSSVQRTAGKTAATVTSALQPPVLSL----- 690
Qy 344 VILHSVGPAAAGCTTAGGGLQTSKPLVTSVANTVTVSSLOPEKRVSGTAVTSLPAV 403
Db 691 -----TOPQVGVGKQGPPLVIO-----QPPKP----- 715
Qy 404 TGETSGAALCLPSVYVFCMDHICKPVI--GTPV-----QIKLAORG 446
Db 716 -----GALIQLNPLQPV-----PYYKPAVLPGTKALSAVSAQAAAAQKNKLEBG 760

RESULT 14
US-09-724-676A-63923
; Sequence 63923, Application US/09724676A
; GENERAL INFORMATION:
```

```
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63923
; LENGTH: 775
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-63923

Query Match
Best Local Similarity 22.6%; Score 507.5; DB 5; Length 775;
Best Local Similarity 32.9%; Pred. No. 1.3e-20;
Matches 157; Conservative 58; Mismatches 121; Indels 141; Gaps 20;

Qy 13 PKVSSG-----PRLAPQIVAVKAPRTTITQPPANIQPPTVILKNSGPI 61
Db 382 PRATTSIGRATLPVLARLP-----PONPTNIQ---NFOLPPGVLVSENGQL 432
Qy 62 LVSPQOVTNR-----AETTSNITSRAVPANDQVTKICTVPSNSQLIKKVAVTPVK 115
Db 433 MI-PQALAQMOQAHAQQTMAAPRATPTSAAPVQISTVOAGTPIIAR-QVTP----- 486
Qy 116 AQIGTVVTVPKPSSVQSAVPTSVVTVPGKPLNTVT--TLKPSLSGASS-----TPS 168
Db 487 -----TTIIQV---SQAQTTVQPSATLQSPGVQPLVGGAAQTASLGTATAVQGTQ 539
Qy 169 NE-PNLKANSAAVQINLSPTMLENKKCKNFLAMLIKILACSGSGSPENQVKKLVEOL 227
Db 540 RTVPGATTTSSATE-----TMENYKCKCKNFLSTLIKILASSGKOSTETAAVKEIYQNL 593
Qy 228 LQAKIAEFTSKLYRELNSPOPIVLPFLKSLPALROLTPDSASFIOOQVQO-----TS 283
Db 594 LQKIEAEFTSKLYRELNSPOPIVLPFLKSLPALROLTPDSAAFIOOQOQPPPPS 653
Qy 284 SDWVATCTTAVTSPVTTTSSOSEKSIIVSGATAPRTVSQTLNPLAGVGAAGV 343
Db 654 Q-----ATTALTAVLSSVQRTAGKTAATVTSALQPPVLSL----- 690
Qy 344 VILHSVGPAAAGCTTAGGGLQTSKPLVTSVANTVTVSSLOPEKRVSGTAVTSLPAV 403
Db 691 -----TOPQVGVGKQGPPLVIO-----QPPKP----- 715
Qy 404 TGETSGAALCLPSVYVFCMDHICKPVI--GTPV-----QIKLAORG 446
Db 716 -----GALIQLNPLQPV-----PYYKPAVLPGTKALSAVSAQAAAAQKNKLEBG 760

RESULT 15
US-09-724-676-63926
; Sequence 63926, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63926
; LENGTH: 807
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-63926

Query Match
Best Local Similarity 22.6%; Score 507.5; DB 5; Length 807;
Best Local Similarity 32.9%; Pred. No. 1.3e-20;
Matches 157; Conservative 58; Mismatches 121; Indels 141; Gaps 20;

Qy 13 PKVSSG-----PRLAPQIVAVKAPRTTITQPPANIQPPTVILKNSGPI 61
```

```

US-09-724-676A-63929
; Sequence 63929, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63920
; LENGTH: 679
; TYPE: PRY
; ORGANISM: Homo sapiens
US-09-724-676A-63929

Query Match
Best Local Similarity 32.9%; Score 507.5; DB 5; Length 679;
Matches 157; Conservative 58; Mismatches 121; Indels 141; Gaps 20;

QY 13 PPKVSSG-----PRLPAPQIVAVKAPNNTTIOFPANLQIPGTVLIRKNSGFLM 61
D 286 PTATSGIRATLPTVLAPRLPOP-----PQNPTNIO--NFOLPPGMVLVRSNGQLL 336
QY 62 LVSPQOQVTR-----AETTSNITSRAVPANPQTVKICVTPNSSSOLIKKVAATVPVKKL 115
D 337 MI-FOQALAQOAHQAHQAPQTTMAPRPRATPSAPPVQISTVQADPGTPIIAR-QVTP----- 390
QY 116 AQIGTTVTVTPKPSVQSVAVPTSVTVTPGKPLNTVT--TLKPSISGASS-----TPS 168
D 391 -----TTTIKQV---SQAOITVQPSATLQRSFCVQPOLVLGGAQATASIGTATVOTGTIPQ 443
QY 169 NE-PNLKAENSAVOINISPTMLENVKCKNFLLMLIKLACSGSGSPMGONVKKLYEOL 227
D 444 RTVPGATTTSSAATE-----TMENVKCKNFLLMLIKLACSGSGOSTETANVVELYQNL 497
QY 228 LDKATEAEFPRKLYVELKSSPOHLPVFLKKSVALROLPPNSQSFIOQCVOQ-----TS 283
D 498 LDGTEADEFTSRILYRELNSSPOPLVPLFKRSLPALROLPPDSAAFTQSSQOQPPPTTS 557
QY 284 SDMVATCTTCTTSPVTVTVSSSSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGV 343
D 558 Q-----ATTALTAVALVSSVQRTAGTATVTSALQPPVLSL-----594
QY 344 VTLHSVGPAAATGTTAGTGLQTSKPLVTSVANTVTVSLQPEKPVVSGTAATLSPAV 403
D 595 -----TOPQOVGVGKQGPPTPLVIO-----QPPKP-----619
QY 404 TFGETSGAATCLPSVKPVPVSCMDHICKPVI--GMPV-----QIKLAOPG 446
D 620 -----GALIQLNPLQPV-----PVYKPAVLPGTKALSAVSAQAAAAQKKKLEPG 664

RESULT 11
US-09-724-676-63920
; Sequence 63920, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63920
; LENGTH: 711
; TYPE: PRY
; ORGANISM: Homo sapiens
US-09-724-676-63920

Query Match
Best Local Similarity 32.9%; Score 507.5; DB 5; Length 711;
Matches 157; Conservative 58; Mismatches 121; Indels 141; Gaps 20;

```

```

QY 13 PPKVSSG-----PRLPAPQIVAVKAPNNTTIOFPANLQIPGTVLIRKNSGFLM 61
D 318 PTATSGIRATLPTVLAPRLPOP-----PQNPTNIO--NFOLPPGMVLVRSNGQLL 368
QY 62 LVSPQOQVTR-----AETTSNITSRAVPANPQTVKICVTPNSSSOLIKKVAATVPVKKL 115
D 369 MI-FOQALAQOAHQAHQAPQTTMAPRPRATPSAPPVQISTVQADPGTPIIAR-QVTP----- 422
QY 116 AQIGTTVTVTPKPSVQSVAVPTSVTVTPGKPLNTVT--TLKPSISGASS-----TPS 168
D 423 -----TTTIKQV---SQAOITVQPSATLQRSFCVQPOLVLGGAQATASIGTATVOTGTIPQ 475
QY 169 NE-PNLKAENSAVOINISPTMLENVKCKNFLLMLIKLACSGSGSPMGONVKKLYEOL 227
D 476 RTVPGATTTSSAATE-----TMENVKCKNFLLMLIKLACSGSGOSTETANVVELYQNL 529
QY 228 LDKATEAEFPRKLYVELKSSPOHLPVFLKKSVALROLPPNSQSFIOQCVOQ-----TS 283
D 530 LDGTEADEFTSRILYRELNSSPOPLVPLFKRSLPALROLPPDSAAFTQSSQOQPPPTTS 589
QY 284 SDMVATCTTCTTSPVTVTVSSSSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGV 343
D 590 Q-----ATTALTAVALVSSVQRTAGTATVTSALQPPVLSL-----626
QY 344 VTLHSVGPAAATGTTAGTGLQTSKPLVTSVANTVTVSLQPEKPVVSGTAATLSPAV 403
D 627 -----TOPQOVGVGKQGPPTPLVIO-----QPPKP-----651
QY 404 TFGETSGAATCLPSVKPVPVSCMDHICKPVI--GMPV-----QIKLAOPG 446
D 652 -----GALIQLNPLQPV-----PVYKPAVLPGTKALSAVSAQAAAAQKKKLEPG 696

RESULT 12
US-09-724-676A-63920
; Sequence 63920, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63920
; LENGTH: 711
; TYPE: PRY
; ORGANISM: Homo sapiens
US-09-724-676A-63920

Query Match
Best Local Similarity 32.9%; Score 507.5; DB 5; Length 711;
Matches 157; Conservative 58; Mismatches 121; Indels 141; Gaps 20;

```

Query Match 23.2%; Score 521; DB 5; Length 837;

Best Local Similarity 34.0%; Pred. No. 2,4e-21;
Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

```

QY 13 PRVSSG-----RLPAPQIVAKAPMTTITIOFPANIQLPBGIVLKSNGPLM 61
DB 414 PRATTSIGIRATLPTVLAPELPQP-----POPNITIQ--NFQLPBGIVLSENGOLL 464
QY 62 LVSPQOQVTR-----AETTSNITSRAVPANQVKTICVTPNSSSOLIKKAVTVPVKUL 115
DB 465 MI-PQOALQMOQAHAQOQTMAPRAPPTSPAPVOISTVQAPGPITIAK-QVTP-----518
QY 116 AQIGTTVTTVPKPSVQSVAVPTSVVTVTPGKPLNTVT--TLKPSLGAAS-----TPS 168
DB 519 -----TTIIKQV--SOAQTTVQPSATLQRSPOVQOLVGAQAQTASLGATVAVQGTFO 571
QY 169 NE-PNKAENSAVQINLSPTMLENKKCKNFIAMLIKILACSSQSPKQONKAKIVEOL 227
DB 572 RIVPGATTTSSATE-----TMENVKCKKNFSTLIKILACSSQSPKQONKAKIVEOL 625
QY 228 LDKAIEAEFTRLKLYELKSSPOPHLVPLFKKSVALLROLPLPSOSFIQOCVQO-----TS 283
DB 626 LDKAIEAEFTSLRYELKSSPOPHLVPLFKKSLPALROLTPDSAAFIOQSOQOPEPPPS 685
QY 284 SDNVIACTTTVTTSPVTTTVSSQSEKSIIVSGATAPRTVSQTLNPLAGVGAAGY 343
DB 686 Q-----ATTALTAVLVSSVQRTAGTAATVTSALQPPVLST-----722
QY 344 VTLHSGPPLAAGCTAGTACGLQTSKPLVTSVANTVTVSLOPEKPVSGTAVTSLPAPV 403
DB 723 -----TQPTQVGVKGQGPPLVLIQ-----QPPKP--GALIRPQV 756
QY 404 TFEETSGAAILCPSPKPVVSCMDHICKPVIQPIQIKL--AQGPVL 449
DB 757 TLTQT-----PMVALRQPH-NRIMLTTPQOIQLNPLQPPVYV 792

```

RESULT 8
US-09-724-676A-63924
Sequence 63924, Application US/09724676A
GENERAL INFORMATION:
APPLICANT: Compugen LTD
TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 Compugen
CURRENT APPLICATION NUMBER: US/09/724,676A
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: PatentIn version 3.2
SEQ ID NO 63924
LENGTH: 837
TYPE: PRV
ORGANISM: Homo sapiens
US-09-724-676A-63924

Query Match 23.2%; Score 521; DB 5; Length 837;
Best Local Similarity 34.0%; Pred. No. 2,4e-21;
Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

```

QY 13 PRVSSG-----RLPAPQIVAKAPMTTITIOFPANIQLPBGIVLKSNGPLM 61
DB 414 PRATTSIGIRATLPTVLAPELPQP-----POPNITIQ--NFQLPBGIVLSENGOLL 464
QY 62 LVSPQOQVTR-----AETTSNITSRAVPANQVKTICVTPNSSSOLIKKAVTVPVKUL 115
DB 465 MI-PQOALQMOQAHAQOQTMAPRAPPTSPAPVOISTVQAPGPITIAK-QVTP-----518
QY 116 AQIGTTVTTVPKPSVQSVAVPTSVVTVTPGKPLNTVT--TLKPSLGAAS-----TPS 168
DB 519 -----TTIIKQV--SOAQTTVQPSATLQRSPOVQOLVGAQAQTASLGATVAVQGTFO 571
QY 169 NE-PNKAENSAVQINLSPTMLENKKCKNFIAMLIKILACSSQSPKQONKAKIVEOL 227
DB 572 RIVPGATTTSSATE-----TMENVKCKKNFSTLIKILACSSQSPKQONKAKIVEOL 625

```

```

QY 228 LDKAIEAEFTRLKLYELKSSPOPHLVPLFKKSVALLROLPLPSOSFIQOCVQO-----TS 283
DB 626 LDKAIEAEFTSLRYELKSSPOPHLVPLFKKSLPALROLTPDSAAFIOQSOQOPEPPPS 685
QY 284 SDNVIACTTTVTTSPVTTTVSSQSEKSIIVSGATAPRTVSQTLNPLAGVGAAGY 343
DB 686 Q-----ATTALTAVLVSSVQRTAGTAATVTSALQPPVLST-----722
QY 344 VTLHSGPPLAAGCTAGTACGLQTSKPLVTSVANTVTVSLOPEKPVSGTAVTSLPAPV 403
DB 723 -----TQPTQVGVKGQGPPLVLIQ-----QPPKP--GALIRPQV 756
QY 404 TFEETSGAAILCPSPKPVVSCMDHICKPVIQPIQIKL--AQGPVL 449
DB 757 TLTQT-----PMVALRQPH-NRIMLTTPQOIQLNPLQPPVYV 792

```

RESULT 9
US-09-724-676-63929
Sequence 63929, Application US/09724676
GENERAL INFORMATION:
APPLICANT: Compugen LTD
TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 Compugen
CURRENT APPLICATION NUMBER: US/09/724,676
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: PatentIn version 3.2
SEQ ID NO 63929
LENGTH: 679
TYPE: PRV
ORGANISM: Homo sapiens
US-09-724-676-63929

Query Match 22.6%; Score 507.5; DB 5; Length 679;
Best Local Similarity 32.9%; Pred. No. 1,1e-20;
Matches 157; Conservative 58; Mismatches 121; Indels 141; Gaps 20;

```

QY 13 PRVSSG-----RLPAPQIVAKAPMTTITIOFPANIQLPBGIVLKSNGPLM 61
DB 286 PRATTSIGIRATLPTVLAPELPQP-----POPNITIQ--NFQLPBGIVLSENGOLL 336
QY 62 LVSPQOQVTR-----AETTSNITSRAVPANQVKTICVTPNSSSOLIKKAVTVPVKUL 115
DB 337 MI-PQOALQMOQAHAQOQTMAPRAPPTSPAPVOISTVQAPGPITIAK-QVTP-----390
QY 116 AQIGTTVTTVPKPSVQSVAVPTSVVTVTPGKPLNTVT--TLKPSLGAAS-----TPS 168
DB 391 -----TTIIKQV--SOAQTTVQPSATLQRSPOVQOLVGAQAQTASLGATVAVQGTFO 443
QY 169 NE-PNKAENSAVQINLSPTMLENKKCKNFIAMLIKILACSSQSPKQONKAKIVEOL 227
DB 444 RIVPGATTTSSATE-----TMENVKCKKNFSTLIKILACSSQSPKQONKAKIVEOL 497
QY 228 LDKAIEAEFTRLKLYELKSSPOPHLVPLFKKSVALLROLPLPSOSFIQOCVQO-----TS 283
DB 498 LDKAIEAEFTSLRYELKSSPOPHLVPLFKKSLPALROLTPDSAAFIOQSOQOPEPPPS 557
QY 284 SDNVIACTTTVTTSPVTTTVSSQSEKSIIVSGATAPRTVSQTLNPLAGVGAAGY 343
DB 558 Q-----ATTALTAVLVSSVQRTAGTAATVTSALQPPVLST-----594
QY 344 VTLHSGPPLAAGCTAGTACGLQTSKPLVTSVANTVTVSLOPEKPVSGTAVTSLPAPV 403
DB 595 -----TQPTQVGVKGQGPPLVLIQ-----QPPKP--GALIRPQV 619
QY 404 TFEETSGAAILCPSPKPVVSCMDHICKPVI--GRPV-----QIKLAQPG 446
DB 620 -----GALIQNLPLQPV-----PVKPAVLPGTKALSAVSAQAAAQKNKLEKPG 664

```


QY 404 TEGTSGAALCLPSVKPVVSPFCMDHICKPVIGTPVOIKL--AOPGPLY 449
 Db 629 TLTQT-----PMVALRQPH-NRIMLTTPQOIQINLPQPVV 664

RESULT 2

US-09-724-676A-63927
 ; Sequence 63927, Application US/09724676A
 ; GENERAL INFORMATION:
 ; APPLICANT: Compugen LTD
 ; TITLE OF INVENTION: Variants of alternative splicing
 ; FILE REFERENCE: 129181.4 Compugen
 ; CURRENT APPLICATION NUMBER: US/09/724, 676A
 ; NUMBER OF SEQ ID NOS: 97222
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 63927
 ; LENGTH: 709
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-724-676A-63927

Query Match 23.2%; Score 521; DB 5; Length 709;
 Best Local Similarity 34.0%; Pred. No. 2.1e-21;
 Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

QY 13 PKVSSG-----PRLPAPQIVAAKAPWTITOPANQILPPTGVILKNSGPI 61
 Db 286 PRATTSGINATLPTVILARLRP-----PQNPITIQ--NFQLPQAVLVSENGQL 336
 QY 62 LVSPQQTVR-----AETTSNITSRPVAVPANDQVTKICTVPNSSQILKVAATPVKKL 115
 Db 337 MI-POQALAQMOQAHAQOPQTMAPRAPATPSAPVQISTVOAGPPIIAR-QVTP----- 390
 QY 116 AQIGTIVTTVPKPSVSVAVPTSVYTPGKPLNTYV--TLKPSGLASS-----TPS 168
 Db 391 ----TTIIKQV--SOAQTTVQPSATLQSPGVQPOLVIGGAAQTASLGTATRAVQGTQ 443
 QY 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFAMLIKILACSGSGSPENGQNVKLEVDL 227
 Db 444 RIVPGATTTSSATE-----TMENVKCKNFSTLIKILACSGSGOSTETRAANVKEVQNL 497
 QY 228 LPAKIAEFTKRLYELKSSPQPHLVPLKKSVALROLLPNSQSFIOCCVQO----TS 283
 Db 498 LDKIAEDFTSRLYRELNSSPQPLVPLKRSIPALROLTPDSAAFIOQSOQPPPTPS 557
 QY 284 SDNVIACTTITVTSVYTTVSSQSEKSIIVSGATPRTVSQVLNPLAGVGAAGV 343
 Db 558 Q-----ATLALTAVLVSSSVQRTAGKTAATVTSALQPPVLSL----- 594
 QY 344 VTLHSVPTAATGTTAGTGLQTSKPLVTSVANTVTVSLQPEKPVSGTAVTSLP 403
 Db 595 -----TQPTQVGVGKQGPPTLVIO-----QPPK-----GALINPPQV 628
 QY 404 TEGTSGAALCLPSVKPVVSPFCMDHICKPVIGTPVOIKL--AOPGPLY 449
 Db 629 TLTQT-----PMVALRQPH-NRIMLTTPQOIQINLPQPVV 664

RESULT 3
 US-09-724-676A-63930
 ; Sequence 63930, Application US/09724676
 ; GENERAL INFORMATION:
 ; APPLICANT: Compugen LTD
 ; TITLE OF INVENTION: Variants of alternative splicing
 ; FILE REFERENCE: 129181.4 Compugen
 ; CURRENT APPLICATION NUMBER: US/09/724, 676
 ; NUMBER OF SEQ ID NOS: 97222
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 63930
 ; LENGTH: 741
 ; TYPE: PRT

; ORGANISM: Homo sapiens
 US-09-724-676A-63930

Query Match 23.2%; Score 521; DB 5; Length 741;
 Best Local Similarity 34.0%; Pred. No. 2.1e-21;
 Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

QY 13 PKVSSG-----PRLPAPQIVAAKAPWTITOPANQILPPTGVILKNSGPI 61
 Db 318 PRATTSGINATLPTVILARLRP-----PQNPITIQ--NFQLPQAVLVSENGQL 368
 QY 62 LVSPQQTVR-----AETTSNITSRPVAVPANDQVTKICTVPNSSQILKVAATPVKKL 115
 Db 369 MI-POQALAQMOQAHAQOPQTMAPRAPATPSAPVQISTVOAGPPIIAR-QVTP----- 422
 QY 116 AQIGTIVTTVPKPSVSVAVPTSVYTPGKPLNTYV--TLKPSGLASS-----TPS 168
 Db 423 ----TTIIKQV--SOAQTTVQPSATLQSPGVQPOLVIGGAAQTASLGTATRAVQGTQ 475
 QY 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFAMLIKILACSGSGSPENGQNVKLEVDL 227
 Db 476 RIVPGATTTSSATE-----TMENVKCKNFSTLIKILACSGSGOSTETRAANVKEVQNL 529
 QY 228 LPAKIAEFTKRLYELKSSPQPHLVPLKKSVALROLLPNSQSFIOCCVQO----TS 283
 Db 530 LDKIAEDFTSRLYRELNSSPQPLVPLKRSIPALROLTPDSAAFIOQSOQPPPTPS 589
 QY 284 SDNVIACTTITVTSVYTTVSSQSEKSIIVSGATPRTVSQVLNPLAGVGAAGV 343
 Db 590 Q-----ATLALTAVLVSSSVQRTAGKTAATVTSALQPPVLSL----- 626
 QY 344 VTLHSVPTAATGTTAGTGLQTSKPLVTSVANTVTVSLQPEKPVSGTAVTSLP 403
 Db 627 -----TQPTQVGVGKQGPPTLVIO-----QPPK-----GALINPPQV 660
 QY 404 TEGTSGAALCLPSVKPVVSPFCMDHICKPVIGTPVOIKL--AOPGPLY 449
 Db 661 TLTQT-----PMVALRQPH-NRIMLTTPQOIQINLPQPVV 696

RESULT 4
 US-09-724-676A-63930
 ; Sequence 63930, Application US/09724676A
 ; GENERAL INFORMATION:
 ; APPLICANT: Compugen LTD
 ; TITLE OF INVENTION: Variants of alternative splicing
 ; FILE REFERENCE: 129181.4 Compugen
 ; CURRENT APPLICATION NUMBER: US/09/724, 676A
 ; NUMBER OF SEQ ID NOS: 97222
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 63930
 ; LENGTH: 741
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-724-676A-63930

Db 595 -----TQPTQVGVGKQGQPTPLVIQ-----QPPKP-----GALIRPPQV 622

This Page Blank (uspto)

```

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
; TITLE OF INVENTION: THEROP
; FILE REFERENCE: C000173
; CURRENT APPLICATION NUMBER: US/60/173,464
; CURRENT FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 30269
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9223
; LENGTH: 921
; TYPE: PRT
; ORGANISM: Drosophila
; US-60-173-464-9223

```

```

Query Match          10.8%; Score 243.5; DB 27; Length 921;
Best Local Similarity 24.8%; Pred. No. 3.6e-10;
Matches 112; Conservative 65; Mismatches 163; Indels 111; Gaps 17;

```

```

QY 25 POIYAVKAPNTTTIOFPANLQLPCTVLKSNGLML-----VSPQOTVTRAEYTS 76
DB 133 PPSPIITLSTLNTGTPA-----LVKTDNGFQLRVGTTGPTTVTQITNTSNNS 184
QY 77 NITSRPVPANPQVTKICTVPSNSQ-----LIKKAVTPYKLAQIGTVVTVTP 127
DB 185 NITSTNNPTTQ-IRLOTVPAASMTNTATSNIIIVNSVSSGVANSOPPHLTQLANQ 243
QY 128 KPSSVQSAVPTSVTVTPGKPLNTVTTLKPSLSGASSTPSNEPMLKAENSAVQINLSP 187
DB 244 APQLPQITQITIPAOOSQOQOVNNVSSAGGATAVSSTTA-----ATT 287
QY 188 TMLENVK-KCKNELMLIKLACSSQSPMGONVKKLYEQLDAKIEAEFTFKLYELK 246
DB 288 TQGGTKCKRKFANLIEL--STREPKPVEKNVRTLQELVNAVNEPEECDRLERLLN 345
QY 247 SSPQHLVPLFKSVVALROL-----LPSQSEFIQ--- 276
DB 346 ASPQCLIGFLKSLPLRQALYTKELVIEGIKPPQHVLAGLSQQLPKIQAOIRPIG 405
QY 277 ---OCVQOTSSDMVIACTTTVTTSPPVYTTVSSSQSEKSIIVSGATAPRTVS---VQT 329
DB 406 PSQTTTIGOTQVRMI--TPNALGTPRPTIGHTTISKOPN--IRLPTAPRLVNTGIRT 460
QY 330 LNPLAGPVGAKAGVYTLHSGPTATGCTAGTGLQTSKPLVTSVANTVTVTSIQPEKP 389
DB 461 QIP-SLOVPGGANIVQIR--GPQHAQLQRTGSVQIRATRRP-----PNSVPTAN----- 506
QY 390 VVSGTAVYLSLPAYTFGETSGAICLPSVKP 420
DB 507 -----KLTAVKVGQTOIKAI--TPSLHP 527

```

```

RESULT 15
; US-60-173-464-20611
; Sequence 20611, Application US/60173464
; GENERAL INFORMATION:
; APPLICANT: LI, Peter W.D.
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
; TITLE OF INVENTION: THEROP
; FILE REFERENCE: C0000173
; CURRENT APPLICATION NUMBER: US/60/173,464
; CURRENT FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 30269
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20611
; LENGTH: 921
; TYPE: PRT
; ORGANISM: Drosophila
; US-60-173-464-20611

```

```

Query Match          10.8%; Score 243.5; DB 27; Length 921;
Best Local Similarity 24.8%; Pred. No. 3.6e-10;
Matches 112; Conservative 65; Mismatches 163; Indels 111; Gaps 17;

```

```

QY 25 POIYAVKAPNTTTIOFPANLQLPCTVLKSNGLML-----VSPQOTVTRAEYTS 76
DB 133 PPSPIITLSTLNTGTPA-----LVKTDNGFQLRVGTTGPTTVTQITNTSNNS 184
QY 77 NITSRPVPANPQVTKICTVPSNSQ-----LIKKAVTPYKLAQIGTVVTVTP 127
DB 185 NITSTNNPTTQ-IRLOTVPAASMTNTATSNIIIVNSVSSGVANSOPPHLTQLANQ 243
QY 128 KPSSVQSAVPTSVTVTPGKPLNTVTTLKPSLSGASSTPSNEPMLKAENSAVQINLSP 187
DB 244 APQLPQITQITIPAOOSQOQOVNNVSSAGGATAVSSTTA-----ATT 287
QY 188 TMLENVK-KCKNELMLIKLACSSQSPMGONVKKLYEQLDAKIEAEFTFKLYELK 246
DB 288 TQGGTKCKRKFANLIEL--STREPKPVEKNVRTLQELVNAVNEPEECDRLERLLN 345
QY 247 SSPQHLVPLFKSVVALROL-----LPSQSEFIQ--- 276
DB 346 ASPQCLIGFLKSLPLRQALYTKELVIEGIKPPQHVLAGLSQQLPKIQAOIRPIG 405
QY 277 ---OCVQOTSSDMVIACTTTVTTSPPVYTTVSSSQSEKSIIVSGATAPRTVS---VQT 329
DB 406 PSQTTTIGOTQVRMI--TPNALGTPRPTIGHTTISKOPN--IRLPTAPRLVNTGIRT 460
QY 330 LNPLAGPVGAKAGVYTLHSGPTATGCTAGTGLQTSKPLVTSVANTVTVTSIQPEKP 389
DB 461 QIP-SLOVPGGANIVQIR--GPQHAQLQRTGSVQIRATRRP-----PNSVPTAN----- 506
QY 390 VVSGTAVYLSLPAYTFGETSGAICLPSVKP 420
DB 507 -----KLTAVKVGQTOIKAI--TPSLHP 527

```

Search completed: February 16, 2003, 22:18:24
Job time : 165.507 secs


```

FILE REFERENCE: 0342/1G548-US1
CURRENT APPLICATION NUMBER: PCT/US01/20592A
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/215,164
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: US 60/224,457
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 146
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 1023
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Corresponds to SEQ ID NO: 87
PCT-US01-20592A-14

```

```

Query Match      22.8%; Score 513; DB 1; Length 1023;
Best Local Similarity 33.4%; Pred. No. 3,4e-32;
Matches 154; Conservative 62; Mismatches 127; Indels 118; Gaps 19;

```

```

OY 13 PPKVSSG-----PRLPAPQIVAVKAPNTTIOFPANLQLPPTGLIKNSGSLM 61
DB 414 PRATSGIRATLPTVLAPRLPQ-----PQNTNIO---NFQLPBGWLVRSNGQL 464
OY 62 LVSPQQTIVTR-----AETSNITSRAVAPNPOTVAKICVPPNSSQLIKKVAATPVKKL 115
DB 465 MI-PQALAQQAQAHQAPQTMAPRATPTSAAPVOISTVQAGFTIAR-QVTP----- 518
OY 116 AQGTIVTVTPKPSVQSAVAVPTSVTPGKPLNTVT--TLKPSLGLASS-----TPS 168
DB 519 -----TTTIKQV---SQAGTIVQPSATLQKSPGQPOLVLGGAQTAISGTAATVQGTGPQ 571
OY 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFLAMLIKLCSSGQSPSEMGQNVKLYEOL 227
DB 572 RTVPGATTTSSAATE-----TMENVKCKNFLSTLIKLCSSGQSTETANVELQNL 625
OY 228 LDKATEEETFRKLYVELKSSQPHLVPLFKKSVVALROLPLNSQSTLOOCVQO-----TS 283
DB 626 LDKATEEEDTSRLKRLNLSPPQYLVPLFKRSLPALROLPLDPSAFTIOSSQOQPPPTPS 685
OY 284 SDMYIATCTTIVTSPVYTTVSSQSEKSIIVSGATAPRTVSQVOTLNPLAGVGARAGV 343
DB 686 Q-----ATTALTAVALVSSVQRTAGKTAATVTSALQPVLSL----- 722
OY 344 VTLHSGPATAAGTGTAGTGLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAATVLSLPAV 403
DB 723 -----TQPTQVGKGQGPPLVIO-----QPKP-----GALIRPPQV 756
OY 404 TFGETSGAATCLPSPKPVVSVFCMDHICKPVIGTPVQIKLQ 444
DB 757 TLTQGT-----PMVALRQPH-NRIMLTTPQOVNLSL 785

```

RESULT 11
US-09-893-519A-14
Sequence 14, Application US/09893519A
GENERAL INFORMATION:
APPLICANT: ANADYS PHARMACEUTICALS, INC.
APPLICANT: THOMPSON, Craig
APPLICANT: MOORE, Jeffrey
APPLICANT: BURMAN, Ed T.
APPLICANT: BRADLEY, John
APPLICANT: DESILVA, Thameera
APPLICANT: HARRIS, Sandra
APPLICANT: KOMARINSKY, Svetlana
APPLICANT: MENDILLO, Marc
APPLICANT: MOORE, Daniel
APPLICANT: MCCOY, Melissa
APPLICANT: SANDERSON, Karen
APPLICANT: HAO, Tariq
APPLICANT: ZHU, Shuhao

```

APPLICANT: LONG, Fan
APPLICANT: DAVIDOV, Eugene
TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
FILE REFERENCE: 0342/1G548-US2
CURRENT APPLICATION NUMBER: US/09/893,519A
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/215,164
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: US 60/224,457
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 146
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 1023
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Corresponds to SEQ ID NO: 87
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Human Genbank/CAA72189
DATABASE ENTRY DATE: 1997-06-25
RELEVANT RESIDUES: (1)..(1023)
US-09-893-519A-14

```

```

Query Match      22.8%; Score 513; DB 22; Length 1023;
Best Local Similarity 33.4%; Pred. No. 3,4e-32;
Matches 154; Conservative 62; Mismatches 127; Indels 118; Gaps 19;

```

```

OY 13 PPKVSSG-----PRLPAPQIVAVKAPNTTIOFPANLQLPPTGLIKNSGSLM 61
DB 414 PRATSGIRATLPTVLAPRLPQ-----PQNTNIO---NFQLPBGWLVRSNGQL 464
OY 62 LVSPQQTIVTR-----AETSNITSRAVAPNPOTVAKICVPPNSSQLIKKVAATPVKKL 115
DB 465 MI-PQALAQQAQAHQAPQTMAPRATPTSAAPVOISTVQAGFTIAR-QVTP----- 518
OY 116 AQGTIVTVTPKPSVQSAVAVPTSVTPGKPLNTVT--TLKPSLGLASS-----TPS 168
DB 519 -----TTTIKQV---SQAGTIVQPSATLQKSPGQPOLVLGGAQTAISGTAATVQGTGPQ 571
OY 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFLAMLIKLCSSGQSPSEMGQNVKLYEOL 227
DB 572 RTVPGATTTSSAATE-----TMENVKCKNFLSTLIKLCSSGQSTETANVELQNL 625
OY 228 LDKATEEETFRKLYVELKSSQPHLVPLFKKSVVALROLPLNSQSTLOOCVQO-----TS 283
DB 626 LDKATEEEDTSRLKRLNLSPPQYLVPLFKRSLPALROLPLDPSAFTIOSSQOQPPPTPS 685
OY 284 SDMYIATCTTIVTSPVYTTVSSQSEKSIIVSGATAPRTVSQVOTLNPLAGVGARAGV 343
DB 686 Q-----ATTALTAVALVSSVQRTAGKTAATVTSALQPVLSL----- 722
OY 344 VTLHSGPATAAGTGTAGTGLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAATVLSLPAV 403
DB 723 -----TQPTQVGKGQGPPLVIO-----QPKP-----GALIRPPQV 756
OY 404 TFGETSGAATCLPSPKPVVSVFCMDHICKPVIGTPVQIKLQ 444
DB 757 TLTQGT-----PMVALRQPH-NRIMLTTPQOVNLSL 785

```

RESULT 12
PCT-US02-30474-3271
Sequence 3271, Application PC/TUS0230474
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyun
APPLICANT: Xue, Aidong J.
APPLICANT: Zhao, Qiang A.
APPLICANT: Wang, Jian-Rui
APPLICANT: Wehrman, Tom

PRIOR APPLICATION NUMBER: PCT/US01/14827
PRIOR FILING DATE: 2001-05-16
NUMBER OF SEQ ID NOS: 3334
SOFTWARE: PL_FL_genes Version 6.0
SEQ ID NO: 1665
LENGTH: 1051
TYPE: PRT
ORGANISM: Homo sapiens
US-60-324-631-1665

Query Match 23.2%; Score 521; DB 27; Length 1051;
Best Local Similarity 34.0%; Pred. No. 7, 8e-33;

Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

QY 13 PPKVSSG-----PRLPAQIVAAKAPNTTITOPANQLPPTGYLIKNSGPI 61
DB 382 PTAATSGIATLPTVLAARLP-----PQNPITIQ---NFQLPQAVLVNSENGL 432
QY 62 LVSPOQVTR-----AETTSNITSRAVPANPQVVKICTVNPSSQLIKKVAATPVKK 115
DB 433 MI-POQALQMOQAHAQPOTTMAPRPAITSPAPVOISTVQAPGPIIAR-QVTP----- 486
QY 116 AQIGTVVTVTPKPSVQSAVPTSVYVTPGKPLNTV--TLKPSLGLSS-----TSS 168
DB 487 ----TTLIKOV---SOAQTVQPSATLORS PGVQPOLVIGAAQVTSLSGTATAVQGTPO 539
QY 169 NE-PNLKENSAAVOINLSPTMLENVKCKNFLAMLIKLAGSGSGSPENGQVKKLEOL 227
DB 540 RYVPGATTTSSATE-----TMENVKCKNFLSTLIKLAGSGSGKQSTETRAANKVLO 593
QY 228 LDKAIEAEFTKRLVELKSSPQPHLVPLKKSVALROLPLNSGFTIOCCVQO---TS 283
DB 594 LDKAIEAEFTKRLVELKSSPQPHLVPLKKSVALROLPLNSGFTIOCCVQO---TS 283
QY 284 SDNVATCTTCTTTSPTVTTSSSEKSIIVSGATFARTVSVOVLNLAGVGAAGV 343
DB 654 Q-----ATTALTAVLVSSVORTAGKTAATVTSALQPPVLSL----- 690
QY 344 VTLHSVPTAATGTTAGTGLQTSKPLVTSVANTVTVSLOPEKPVSGTAVTSLPAY 403
DB 691 -----TOPQVGVKGOGQPTPLVIO-----QPKP-----GALLRPQV 724
QY 404 TGEISGAALCLPSVKPVVSPFCMDHICKPVIGTPVOIKL--AQPGVYL 449
DB 725 TLTQT-----PMVALRQPH-NRIMLTTPQOIQLNPLQPVVV 760

RESULT 9
PCT-US01-20592-14
Sequence 14, Application PC/TUS0120592

GENERAL INFORMATION:
APPLICANT: ANADYS PHARMACEUTICALS, INC.

APPLICANT: THOMPSON, Craig
APPLICANT: MOORE, Jeffrey
APPLICANT: BURMAN, Ed T.
APPLICANT: BRADLEY, John
APPLICANT: DESILVA, Thamara
APPLICANT: HARRIS, Sandra
APPLICANT: KOMARNITSKY, Svetlana
APPLICANT: MENDILLO, Marc
APPLICANT: MOORE, Daniel
APPLICANT: MCCOY, Melissa
APPLICANT: SANDERSON, Karen
APPLICANT: HAO, Tariq
APPLICANT: ZHU, Shuhao
APPLICANT: LONG, Fan
APPLICANT: DAVIDOV, Eugene
TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
FILE REFERENCE: 0342/16548-US1
CURRENT APPLICATION NUMBER: PCT/US01/20592
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/215,164
PRIOR FILING DATE: 2000-06-29

PRIOR APPLICATION NUMBER: US 60/224,457
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 146
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 14
LENGTH: 1023
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: human genbank accession #: CAA72189
NAME/KEY: misc.feature
OTHER INFORMATION: Corresponds to SEQ ID NO: 87
PCT-US01-20592-14

Query Match 22.8%; Score 513; DB 1; Length 1023;
Best Local Similarity 33.4%; Pred. No. 3, 4e-32;

Matches 154; Conservative 62; Mismatches 127; Indels 118; Gaps 19;

QY 13 PPKVSSG-----PRLPAQIVAAKAPNTTITOPANQLPPTGYLIKNSGPI 61
DB 414 PTAATSGIATLPTVLAARLP-----PQNPITIQ---NFQLPQAVLVNSENGL 464
QY 62 LVSPOQVTR-----AETTSNITSRAVPANPQVVKICTVNPSSQLIKKVAATPVKK 115
DB 465 MI-POQALQMOQAHAQPOTTMAPRPAITSPAPVOISTVQAPGPIIAR-QVTP----- 518
QY 116 AQIGTVVTVTPKPSVQSAVPTSVYVTPGKPLNTV--TLKPSLGLSS-----TSS 168
DB 519 ----TTLIKOV---SOAQTVQPSATLORS PGVQPOLVIGAAQVTSLSGTATAVQGTPO 571
QY 169 NE-PNLKENSAAVOINLSPTMLENVKCKNFLAMLIKLAGSGSGSPENGQVKKLEOL 227
DB 572 RYVPGATTTSSATE-----TMENVKCKNFLSTLIKLAGSGSGKQSTETRAANKVLO 625
QY 228 LDKAIEAEFTKRLVELKSSPQPHLVPLKKSVALROLPLNSGFTIOCCVQO---TS 283
DB 626 LDKAIEAEFTKRLVELKSSPQPHLVPLKKSVALROLPLNSGFTIOCCVQO---TS 283
QY 284 SDNVATCTTCTTTSPTVTTSSSEKSIIVSGATFARTVSVOVLNLAGVGAAGV 343
DB 686 Q-----ATTALTAVLVSSVORTAGKTAATVTSALQPPVLSL----- 722
QY 344 VTLHSVPTAATGTTAGTGLQTSKPLVTSVANTVTVSLOPEKPVSGTAVTSLPAY 403
DB 723 -----TOPQVGVKGOGQPTPLVIO-----QPKP-----GALLRPQV 756
QY 404 TGEISGAALCLPSVKPVVSPFCMDHICKPVIGTPVOIKLAQ 444
DB 757 TLTQT-----PMVALRQPH-NRIMLTTPQOVNLSE 785

RESULT 10
PCT-US01-20592A-14
Sequence 14, Application PC/TUS0120592A

GENERAL INFORMATION:
APPLICANT: ANADYS PHARMACEUTICALS, INC.

APPLICANT: THOMPSON, Craig
APPLICANT: MOORE, Jeffrey
APPLICANT: BURMAN, Ed T.
APPLICANT: BRADLEY, John
APPLICANT: DESILVA, Thamara
APPLICANT: HARRIS, Sandra
APPLICANT: KOMARNITSKY, Svetlana
APPLICANT: MENDILLO, Marc
APPLICANT: MOORE, Daniel
APPLICANT: MCCOY, Melissa
APPLICANT: SANDERSON, Karen
APPLICANT: HAO, Tariq
APPLICANT: ZHU, Shuhao
APPLICANT: LONG, Fan
APPLICANT: DAVIDOV, Eugene
TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE

```

: APPLICANT: Ghosh, Malabika
: APPLICANT: Wang, Dunrul
: APPLICANT: Ma, Yungqing
: APPLICANT: Asundi, Vinod
: APPLICANT: Wang, Zhiwei
: APPLICANT: Meng, Gezhi
: APPLICANT: Haley-Vicente, Dana
: APPLICANT: Drmanac, Radoje T
: TITLE OF INVENTION: Novel Nucleic Acids and
: FILE OF INVENTION: Polypeptides
: FILE REFERENCE: 810CIP PCT
: CURRENT APPLICATION NUMBER: PCT/US02/30474
: CURRENT FILING DATE: 2002-09-24
: PRIOR APPLICATION NUMBER: US 60/324,631
: PRIOR FILING DATE: 2001-09-24
: PRIOR APPLICATION NUMBER: US 09/488,725
: PRIOR FILING DATE: 2000-01-21
: PRIOR APPLICATION NUMBER: US 09/552,317
: PRIOR FILING DATE: 2000-04-25
: PRIOR APPLICATION NUMBER: PCT/US00/35017
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: US 09/491,404
: PRIOR FILING DATE: 2000-01-25
: PRIOR APPLICATION NUMBER: PCT/US01/02623
: PRIOR FILING DATE: 2001-01-25
: PRIOR APPLICATION NUMBER: US 09/496,914
: PRIOR FILING DATE: 2000-02-03
: PRIOR APPLICATION NUMBER: US 09/560,875
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: PCT/US01/03800
: PRIOR FILING DATE: 2001-02-05
: PRIOR APPLICATION NUMBER: US 09/515,126
: PRIOR FILING DATE: 2000-02-28
: Remaining Prior Application data removed - See file wrapper or PALM.
: NUMBER OF SEQ ID NOS: 3476
: SOFTWARE: PL_FL-genes Version 6.0
: SEQ ID NO 1660
: LENGTH: 1051
: TYPE: PRT
: ORGANISM: Homo sapiens
PCT-0502-30474-1660

Query Match      23.2% Score 521; DB 1: Length 1051;
Best Local Similarity 34.0%; Pred. No. 7.8e-33;
Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

QY 13 PPKVSSG-----PRLAPQIVAVKAPNTTIIQFPANQLPPTVLIKNSGFLM 61
DB 382 PLATTSIGIRATLPPTVLAPRLPQP-----PQNTNIO---NFQLPQMTLVBSNGQLL 432
QY 62 LVSPQQTIVTR-----AETTSNITSRPAVPANPQTVKICIVPNSSQLIKKVAVTVPKKL 115
DB 433 MI-PQQAIAQMAQAHAPQTTMAPRPATPTSAAPVQISTVQACPTPIAR-QVTP----- 486
QY 116 AQIGTIVTVTPKPSVOSAVPVSVTYMPGRKLNTYT--TLKPPSSIGASS-----TPS 168
DB 487 ----TTIKOY--SQAQTVQPSATLQSRSPGVOPQLVLGGAQATASIGATVAVQTGTP 539
QY 169 NE-PNLAKENSAVAQVIMLSPTLLENVKKCKNFLLMLIKLACSGSQSPMGONVKVLEOL 227
DB 540 RTVGAGATTSSAATE-----TMENVKKCKNFLLSTLIKILASSGSGSTETANVKEIVNLT 593
QY 228 LDAKIEAEETFRKLIYVELKSSPOPHLVLPFLKKSVALROLPLNSQSFIOQCVOO-----TS 283
DB 594 LDGKIEAEDEFTSRLRYRELNSPQPYLVLPFLKRSPLALROLPPDSAAFIQGSQOQPPPTS 653
QY 284 SDWIAICTTIVTSPVTTVSSQSEKSIYSGATAPRVSQTLNPLAGPVGARAGV 343
DB 654 Q-----ATTALAVAVLSSSVQFTAGKTATVTSALQPVLSL----- 690
QY 344 VTLHSVGPATAGTGTAGTGLQTSKPLVTSVANTVTTVSLQPKPVSVGTAVTLSLAPV 403
DB 691 -----TQPTQYGVGKQSGOPTPLVIO-----QPKP-----GALIRPPQV 724
```

```

QY 404 TFGENGAICLPSPYKPVVSCWMDICKPVIGTPIVQIKL--AQGCPVL 449
DB 725 TLTQF-----PMVALRQPH-NRIMMLTTPOQIOLNPLQPVV 760

RESULT 8
US-60-324-631-1665
: Sequence 1665, Application US/60324631
: GENERAL INFORMATION:
: APPLICANT: Tang, Y. Tom
: APPLICANT: Zhang, Jie
: APPLICANT: Ren, Feiyun
: APPLICANT: Xue, Aiding J.
: APPLICANT: Zhao, Qing A.
: APPLICANT: Wang, Jian-Rui
: APPLICANT: Wehrman, Tom
: APPLICANT: Zhou, Ping
: APPLICANT: Ghosh, Malabika
: APPLICANT: Wang, Dunrul
: APPLICANT: Ma, Yungqing
: APPLICANT: Asundi, Vinod
: APPLICANT: Wang, Zhiwei
: APPLICANT: Meng, Gezhi
: APPLICANT: Haley-Vicente, Dana
: APPLICANT: Drmanac, Radoje T
: TITLE OF INVENTION: Novel Nucleic Acids and
: FILE OF INVENTION: Polypeptides
: FILE REFERENCE: 810
: CURRENT APPLICATION NUMBER: US/60/324,631
: CURRENT FILING DATE: 2001-09-24
: PRIOR APPLICATION NUMBER: US 09/488,725
: PRIOR FILING DATE: 2000-01-21
: PRIOR APPLICATION NUMBER: US 09/552,317
: PRIOR FILING DATE: 2000-04-25
: PRIOR APPLICATION NUMBER: PCT/US00/35017
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: US 09/491,404
: PRIOR FILING DATE: 2000-01-25
: PRIOR APPLICATION NUMBER: PCT/US01/02623
: PRIOR FILING DATE: 2001-01-25
: PRIOR APPLICATION NUMBER: US 09/496,914
: PRIOR FILING DATE: 2000-02-03
: PRIOR APPLICATION NUMBER: US 09/560,875
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: PCT/US01/03800
: PRIOR FILING DATE: 2001-02-05
: PRIOR APPLICATION NUMBER: US 09/515,126
: PRIOR FILING DATE: 2000-02-28
: PRIOR APPLICATION NUMBER: US 09/577,409
: PRIOR FILING DATE: 2000-05-18
: PRIOR APPLICATION NUMBER: PCT/US01/04927
: PRIOR FILING DATE: 2001-02-26
: PRIOR APPLICATION NUMBER: US 09/519,705
: PRIOR FILING DATE: 2000-03-07
: PRIOR APPLICATION NUMBER: US 09/574,454
: PRIOR FILING DATE: 2000-05-19
: PRIOR APPLICATION NUMBER: PCT/US01/04941
: PRIOR FILING DATE: 2001-03-05
: PRIOR APPLICATION NUMBER: US 09/540,217
: PRIOR FILING DATE: 2000-03-31
: PRIOR APPLICATION NUMBER: US 09/649,167
: PRIOR FILING DATE: 2000-08-23
: PRIOR APPLICATION NUMBER: PCT/US01/08631
: PRIOR FILING DATE: 2001-03-30
: PRIOR APPLICATION NUMBER: US 09/552,929
: PRIOR FILING DATE: 2000-04-18
: PRIOR APPLICATION NUMBER: US 09/770,160
: PRIOR FILING DATE: 2001-01-26
: PRIOR APPLICATION NUMBER: PCT/US01/08656
: PRIOR FILING DATE: 2001-04-18
: PRIOR APPLICATION NUMBER: US 09/577,408
: PRIOR FILING DATE: 2000-05-18
```



```

: PRIOR APPLICATION NUMBER: US 60/328,185
: PRIOR FILING DATE: 2001-10-05
: NUMBER OF SEQ ID NOS: 66
: SOFTWARE: PERL Program
: SEQ ID NO 25
: LENGTH: 865
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: OTHER INFORMATION: Incyte ID No: 4398735CD1
PCT-US02-25829-25

```

```

Query Match          99.5%; Score 2237.5; DB 1; Length 865;
Best Local Similarity 99.3%; Pred. No. 1.1e-173;
Matches 452; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

```

```

QY 1 GTTATKAPVAPPKVSSGRLPAPQIVAKAPMTTITOPANQLP---GTVLINSNS 57
Db 67 GTTATKAPVAPPKVSSGRLPAPQIVAKAPMTTITOPANQLP---GTVLINSNS 126
QY 58 GPLMVSPOQTVTRAETTSNITSRPAPANPQTVKICTVNSSQLIKKAVTPVKKLAQ 117
Db 127 GPLMVSPOQTVTRAETTSNITSRPAPANPQTVKICTVNSSQLIKKAVTPVKKLAQ 186
QY 118 IGTTVTVTRKPSVQSVANPTSVYVTPGKPLNTVTTLKPSLGASSTPSNEPNLKAEN 177
Db 187 IGTTVTVTRKPSVQSVANPTSVYVTPGKPLNTVTTLKPSLGASSTPSNEPNLKAEN 246
QY 178 SAANOILSPMLLENVKKCNFLMLIKLACSGSPKMGONVKKLVEQLDAIEAEF 237
Db 247 SAANOILSPMLLENVKKCNFLMLIKLACSGSPKMGONVKKLVEQLDAIEAEF 306
QY 238 TRRLVELKSSPQHLVPLFKSVVALROLLPNSQSFIOQCVOQTSSDMVATCTTVT 297
Db 307 TRRLVELKSSPQHLVPLFKSVVALROLLPNSQSFIOQCVOQTSSDMVATCTTVT 366
QY 298 SPVTTTVSSQSEKSLIISGATAPRTVSQVTLNPLAGPVGAKGVYTLHSVGTAAATG 357
Db 367 SPVTTTVSSQSEKSLIISGATAPRTVSQVTLNPLAGPVGAKGVYTLHSVGTAAATG 426
QY 358 TTAGTGLQTSKPLVTSVANTVTTVSLOPEKPVSGTAVTLSPAVTFEGESGAATCLPS 417
Db 427 TTAGTGLQTSKPLVTSVANTVTTVSLOPEKPVSGTAVTLSPAVTFEGESGAATCLPS 486
QY 418 VKPVSEFCMDHICKPVIGTGVQIKLAQPGVLSQP 452
Db 487 VKPVSEFCMDHICKPVIGTGVQIKLAQPGVLSQP 521

```

```

RESULT 4
PCT-US02-29964-410
: Sequence 410, Application PC/TUS0229964
: GENERAL INFORMATION:
: APPLICANT: Tang, Y. Tom
: APPLICANT: Ren, Feiyun
: APPLICANT: Zhang, Jie
: APPLICANT: Zhao, Qing A.
: APPLICANT: Wang, Jian-Rui
: APPLICANT: Xue, Aifeng J.
: APPLICANT: Wang, Dunrui
: APPLICANT: Ghosh, Malabika
: APPLICANT: Asundi, Vinod
: APPLICANT: Wehrman, Tom
: APPLICANT: Zhou, Ping
: APPLICANT: Goodrich, Ryle W.
: APPLICANT: Meng, Gezhi
: APPLICANT: Haley-Vicente, Dana
: APPLICANT: Drmanac, Radoje T
: TITLE OF INVENTION: Novel Nucleic Acids and
: FILE REFERENCE: 809ACIP PCT
: CURRENT APPLICATION NUMBER: PCT/US02/29964

```

```

: CURRENT FILING DATE: 2002-09-19
: PRIOR APPLICATION NUMBER: US 60/323,739
: PRIOR FILING DATE: 2001-09-19
: PRIOR APPLICATION NUMBER: US 09/488,725
: PRIOR FILING DATE: 2000-01-21
: PRIOR APPLICATION NUMBER: US 09/552,317
: PRIOR FILING DATE: 2000-04-25
: PRIOR APPLICATION NUMBER: PCT/US00/35017
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: US 09/491,404
: PRIOR FILING DATE: 2000-01-25
: PRIOR APPLICATION NUMBER: PCT/US01/02623
: PRIOR FILING DATE: 2001-01-25
: PRIOR APPLICATION NUMBER: US 09/496,914
: PRIOR FILING DATE: 2000-02-03
: PRIOR APPLICATION NUMBER: US 09/560,875
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: PCT/US01/03800
: PRIOR FILING DATE: 2001-02-05
: PRIOR APPLICATION NUMBER: US 09/515,126
: Remaining Prior Application data removed - See File Wrapper or PALM.
: SOFTWARE: PL_FL_genes Version 6.0
: SEQ ID NO 410
: LENGTH: 843
: TYPE: PRT
: ORGANISM: Homo sapiens
PCT-US02-29964-410

```

```

Query Match          97.1%; Score 2184; DB 1; Length 843;
Best Local Similarity 96.5%; Pred. No. 2.6e-169;
Matches 445; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

QY 1 GTTATKAPVAPPKVSSGRLPAPQIVAKAPMTTITOPANQLP---GTVLINSNS 60
Db 43 GTTATKAPVAPPKVSSGRLPAPQIVAKAPMTTITOPANQLP---GTVLINSNS 102
QY 61 MLVSPQQTVAETTSNITSRPAPANPQTVKICTVNSSQLIKKAVTPVKKLAQIGT 120
Db 103 MLVSPQQTVAETTSNITSRPAPANPQTVKICTVNSSQLIKKAVTPVKKLAQIGT 162
QY 121 TVTTVTRKPSVQSVANPTSVYVTPGKPLNTVTTLKPSLGASSTPSNEPNLKAENSA 180
Db 163 TVTTVTRKPSVQSVANPTSVYVTPGKPLNTVTTLKPSLGASSTPSNEPNLKAENSA 222
QY 181 VOINLSPMLLENVKKCNFLMLIKLACSGSPKMGONVKKLVEQLDAIEAEFTRK 240
Db 223 VOINLSPMLLENVKKCNFLMLIKLACSGSPKMGONVKKLVEQLDAIEAEFTRK 282
QY 241 LVLELKSSPQHLVPLFKSVVALROLLPNSQSFIOQCVOQTSSDMVATCTTVTSPV 300
Db 283 LVLELKSSPQHLVPLFKSVVALROLLPNSQSFIOQCVOQTSSDMVATCTTVTSPV 342
QY 301 VTTTVSSQSEKSLIISGATAPRTVSQVTLNPLAGPVGAKGVYTLHSVGTAAATGTTA 360
Db 343 VTTTVSSQSEKSLIISGATAPRTVSQVTLNPLAGPVGAKGVYTLHSVGTAAATGTTA 402
QY 361 GTGLQTSKPLVTSVANTVTTVSLOPEKPVSGTAVTLSPAVTFEGESGAATCLPSVK 420
Db 403 GTGLQTSKPLVTSVANTVTTVSLOPEKPVSGTAVTLSPAVTFEGESGAATCLPSVK 462
QY 421 VVSEFCMDHICKPVIGTGVQIKLAQPGVLSQP 452
Db 463 VVSEFCMDHICKPVIGTGVQIKLAQPGVLSQP 494

```

```

RESULT 5
US-60-243-468-1271
: Sequence 1271, Application US/60243468
: GENERAL INFORMATION:
: APPLICANT: Beasley, Ellen
: TITLE OF INVENTION: ISOLATED HUMAN DRUG TARGET PROTEINS,

```

Qy	61	MLVSPQOQVTAETTSNTTSPAPANPOFKICTVNSSOLIKKVAYPVKKLOIGT	120
Db	61	MLVSPQOQVTAETTSNTTSPAPANPOFKICTVNSSOLIKKVAYPVKKLOIGT	120
Qy	121	TVVTVTPKPSVQSAVAEPTSVVTVTPKPLNTVTTTLKPSISGASPTSPNPNLKAENSA	180
Db	121	TVVTVTPKPSVQSAVAEPTSVVTVTPKPLNTVTTTLKPSISGASPTSPNPNLKAENSA	180
Qy	181	VOINISPLMELNVKCKCNFLMLIKLACSGSGSPBMCQNNKKLVEOLLDAKIEAEFTPK	240
Db	181	VOINISPLMELNVKCKCNFLMLIKLACSGSGSPBMCQNNKKLVEOLLDAKIEAEFTPK	240
Qy	241	LYVELKSSPOEHLVPELKSSVVALROLLPNNSQSTIOOCVOOTSSDMVIAICTTIVTTS	300
Db	241	LYVELKSSPOEHLVPELKSSVVALROLLPNNSQSTIOOCVOOTSSDMVIAICTTIVTTS	300
Qy	301	VTTTVSSQSEKSIIVSGATAPRTVSVQTLNPLAGPVGAGAKGVTTLHSVPTAATGTTA	360
Db	301	VTTTVSSQSEKSIIVSGATAPRTVSVQTLNPLAGPVGAGAKGVTTLHSVPTAATGTTA	360
Qy	361	GTGLLOTSKPLVTVSVANTVTTVSLQPEKPVVSGTAVTLSLPAVTFGETSGAALCLP	420
Db	361	GTGLLOTSKPLVTVSVANTVTTVSLQPEKPVVSGTAVTLSLPAVTFGETSGAALCLP	420
Qy	421	VVSCMDHICKPVIGTPVQIKLAPGCVLSLP	452
Db	421	VVSCMDHICKPVIGTPVQIKLAPGCVLSLP	452

RESULT 2
US-09-763-909-2
; Sequence 2, Application US/09763909

```

: APPLICANT: Dikstein, Riva
: APPLICANT: Yamit-Hezi, Avai
: TITLE OF INVENTION: A TRANSCRIPTION FACTOR TRIFID SUBUNIT,
: THEREFOR AND
: TITLE OF INVENTION: TAF1105, POLYPEPTIDES, DNA ENCODING
: TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS
: FILE REFERENCE: 13005/002001
: CURRENT APPLICATION NUMBER: US/09/763,909
: CURRENT FILING DATE: 2001-02-26
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 852
: TYPE: prt
: ORGANISM: Homo sapiens
: US-09-763-909-2

```

Query Match	100.0%	Score 2249	DB 21	Length 852
Best Local Similarity	100.0%	Pred. No. 1.2e+174		
Matches 452	0	Mismatches 0	Indels 0	Gaps 0

[illegible]

Qy	301	VTTTAVSSSSQSEKSIIVS	ATAPRVS	QVLTINPL	PLAGP	AGAKAGV	TTHTSV	PTATGGTTA	360
Db	301	VTTTAVSSSSQSEKSIIVS	GATAPRVS	QVLTINPL	PLAGP	AGAKAGV	TTHTSV	PTATGGTTA	360
Qy	361	GTGLQTSKPLPVT	SVANTVT	TVLSQ	PEKPV	SGNAV	LTSLPAT	FGPESAAICLP	SVKP
Db	361	GTGLQTSKPLPVT	SVANTVT	TVLSQ	PEKPV	SGNAV	LTSLPAT	FGPESAAICLP	SVKP
Qy	421	VVSECDHICKP	VIETP	QVQIKL	APG	PVLS	LSOP		452
Db	421	VVSECDHICKP	VIETP	QVQIKL	APG	PVLS	LSOP		452

RESULT 3
PCT-US02-25829-25

Sequence 25, Application PC/TUS0225829

GENERAL INFORMATION:

APPLICANT: INCYTE GENOMICS, INC.

APPLICANT: BAROSSO, Ines

APPLICANT: BAUGHN, Mariah R.

APPLICANT: BECHA, Shanya D.

APPLICANT: BLAKE, Julie J.

APPLICANT: BOROMSKY, Mark L.

APPLICANT: BURFORD, Neil

APPLICANT: DUGGAN, Brendan M.

APPLICANT: ELLIOTT, Ellick S.

APPLICANT: EMERLING, Brooke M.

APPLICANT: FORSTHEE, Ian J.

APPLICANT: GIETZEN, Kimberly J.

APPLICANT: GORVAD, Ann E.

APPLICANT: GRIFFIN, Jennifer A.

APPLICANT: HAFALIA, April J.A.

APPLICANT: HONCHELL, Cynthia D.

APPLICANT: ISON, Craig H.

APPLICANT: KHAN, Farah A.

APPLICANT: TAL, Preeti G.

APPLICANT: LEE, Ernestine A.

APPLICANT: LEE, Sally

APPLICANT: LEE, Soo Yeun

APPLICANT: LI, Joana X.

APPLICANT: LU, Dyrung Aina M.

APPLICANT: LU, Yan

APPLICANT: LEHR-MASON, Patricia M.

APPLICANT: NGUYEN, Daniel B.

APPLICANT: RAMKUMAR, Jayalaxmi

APPLICANT: SERRAQUE, William W.

APPLICANT: TANG, Y. Tom

APPLICANT: THANGAVELU, Kavitha

APPLICANT: THORNTON, Michael

APPLICANT: TRAN,uyen K.

APPLICANT: WALIA, Nairidget R.

APPLICANT: WARREN, Birdiget A.

APPLICANT: XU, Yuming

APPLICANT: YAO, Monique G.

APPLICANT: YEE, Henry

APPLICANT: YEE, Hublin

APPLICANT: ZEBARACADIAN, Yeganeh

TITLE OF INVENTION: NUCLEIC ACID ASSOCIATED PROTEINS

FILE REFERENCE: PF-1146 PCT

CURRENT APPLICATION NUMBER: PCT/US02/25829

CURRENT FILING DATE: 2002-08-14

PRIOR APPLICATION NUMBER: US 60/313,111

PRIOR FILING DATE: 2001-08-17

PRIOR APPLICATION NUMBER: US 60/314,682

PRIOR FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: US 60/314,756

PRIOR FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: US 60/315,105

PRIOR FILING DATE: 2001-08-27

PRIOR APPLICATION NUMBER: US 60/316,751

PRIOR FILING DATE: 2001-08-31

PRIOR APPLICATION NUMBER: US 60/316,856

PRIOR FILING DATE: 2001-08-31

This Page Blank (uspto)

QY 8 AAVSAPKVSQGRPLPAPQIVAVKAPNT-----TTIOFPANLOLPPGTVL 53
 DB 590 APTTPREPAPTPKEPAPTPKETAATTPKGTAPTLKEPAPTPKRAPKELAPTT-----646
 QY 54 KNSGSLMIVSQOYVTRAEITSNITSRAVAVANPOTVACIVPNSSOLIKKAVANTPVK 113
 DB 647 --TKEPTSTSDKPAPTTPKGTAPTPKEPAPTPKGTAPTPKGTAPTPKGTAPTPK 704
 QY 114 KLAQIGTVVTVVPKPSVQSVAVPTS---VTVPGKPLNTVTLKPPSLGASSTPSNE 170
 DB 705 KAPKELAPTTTKGPTSTSDKPAPTTPKETAATTPKETAAPT--TPKKP-----APTPEP 759
 QY 171 PUKAENSAVOJINLSPMLLENVKKCNFLMLIKLACSGSOSPENGQVKKLVEQLDA 230
 DB 760 PPTTSEVSPPTTK--EPTTIH-----KSP-----782
 QY 231 KIEAEFTKRLVELKSSQPHVLPPLKSVVALROLNLSQSFIOOCVOQTSNDVIAT 290
 DB 783 ---DEST---PELSAETPP-----KALENSKEGCVPT 809
 QY 291 CTTVTVSPVTTVSSQSEKSIISGATAPRTVSQTLNPLAGVAGKAGVTLHSVG 350
 DB 810 TTPATKREMTTAKDKTERDL---RTTPTTT-----AAPKMKETAATTEKTT 858
 QY 351 PTAATGCTTAGTGL-----LQTSKPLVTSVANTVTVSL--QPERK 388
 DB 859 ESKITATTTQVSTTQDTTPFKITTLKTTTLAPKVTITTKTITTEIMNKPEE 912

RESULT 13
 US-09-801-368-110
 ; Sequence 110, Application US/09801368
 ; Patent No. US20020128250A1
 ; GENERAL INFORMATION:

APPLICANT: Busby, Robert
 APPLICANT: Call, Brian
 APPLICANT: Hecht, Peter
 APPLICANT: Holtzman, Doug
 APPLICANT: Madden, Kevin
 APPLICANT: Maxon, Mary
 APPLICANT: Moline, Todd
 APPLICANT: No. US20020128250A1man, Thea
 APPLICANT: Royer, John
 APPLICANT: Salama, Sofie
 APPLICANT: Sherman, Amir
 APPLICANT: Silva, Jeff
 APPLICANT: Summers, Eric
 TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
 FILE REFERENCE: 109272.147
 CURRENT APPLICATION NUMBER: US/09/801,368
 CURRENT FILING DATE: 2001-03-07
 PRIOR APPLICATION NUMBER: US 09/487,558
 PRIOR FILING DATE: 2000-01-19
 PRIOR APPLICATION NUMBER: US 60/160,587
 PRIOR FILING DATE: 1999-10-20
 NUMBER OF SEQ ID NOS: 440
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 110
 LENGTH: 1075
 TYPE: PRT
 ORGANISM: Saccharomyces cerevisiae
 US-09-801-368-110

Query Match 6.9%; Score 156; DB 10; Length 1075;
 Best Local Similarity 22.4%; Pred. No. 0 005;

Matches 101; Conservative 63; Mismatches 170; Indels 116; Gaps 21;

QY 2 TLTVAVAPYAPKVSQGRPLPAPQIVAVKAPN-----TTIOFPANLOLPPGTVLKSN 56
 DB 427 TEVTITGNGQPTDET-----VIVIRPTSEGLITTTTE-----PWTGFTST 470
 QY 57 SGPMLV---SQQT-----VTRAEITSNITSRAVP-----ANPQIVKICIVPNSSOLI 104

DB 471 STEMTVTGNGQPTDEVIIVIRPTSEGLISTTEPMTGTETSTETVTLTGNGQPT 530
 QY 105 KKVAV---TPVKKLAQIGTVVTVVPKPSVQSVAVANPISVTVTPKPLN--TTTLKPS 160
 DB 531 DEVIIVIRPTSE-----GLITTTTEPMTGTETSTETVTLTGNGQPTDEVIIVIRPT 586
 QY 161 LGASSTPSNEPNKAENSAVOI-----NLSPMLLENVKKCNFLMLIK--LACSGS 211
 DB 587 SEGLIRRTTEPMTGTETSTETVTLTGNGQPT-----DEVIIVIRPTTLISS 637
 QY 212 QSPEMQNVKKLVEQLDAKIEAEFTKRLVELKSSQPHVLPF-----LKRSV- 262
 DB 638 LSSSSQIATSI-----TSSRPILTFPYSNGTSVISSSVIS 674
 QY 263 --ALROLPPNSQSFIOOCY-----QQTSDMVIATCTVTVTSPTVTVTVSSQ 309
 DB 675 SSVTSSLVTSSTSSSTSSSTSTSTSTSTSTSTSTSTSTSTSTSTSTSSSEKTSASSSS 734
 QY 310 SEKSIIVSGATAPRTVSQTLNPLAGVAGKAGVTLHSVGPTAATGTTAGTGLQTSK 369
 DB 735 SSSSISSESPKSP--TSSSSLPV---TSATTCQETASSL--PRAITTKTSE-----QTL 784
 QY 370 PLVTSVANTVTVSLOPEKPVSGTAVTUS 399
 DB 785 VTVTSCESHVCTESI--SSAIVSTATVTVS 812

RESULT 14
 US-10-124-557-104
 ; Sequence 104, Application US/10124557
 ; Patent No. US20020137894A1
 ; GENERAL INFORMATION:

APPLICANT: Turner, Katherine
 Clark, Stephen C.
 Jacobs, Kenneth
 Hewick, Rodney M.
 Gesner, Thomas G.
 TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 CambridgePark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/124,557
 FILING DATE: 16-Apr-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Cseert, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)876-1170
 TELEFAX: (617)876-5851
 INFORMATION FOR SEQ ID NO: 104:

Db 832 ESKITATTQVSTTODTTPFKITTKTTLAKVTTTKTITTEIMKPEE 885

RESULT 11

US-10-124-557-74
; Sequence 74, Application US/10124557
; Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cserr, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 74:

SEQUENCE CHARACTERISTICS:

LENGTH: 1038 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 74:

US-10-124-557-74

Query Match 6.9%; Score 156; DB 12; Length 1038;
Best Local Similarity 20.8%; Pred. No. 0.0048;
Matches 86; Conservative 40; Mismatches 164; Indels 124; Gaps 14;

QY 8 APVSAPKVSGRRLAPQVAVAKAPNT-----TTIQPPANLQLPFGVLI 53
Db 547 APTTPKEPAPPTPEPAPTPKEPAPTPKGTAPTTLKEPAPTPPKKPAKELAPTT--- 603
QY 54 KSNAGPLMLVSPQOTVRAETSNITSRPAVPANPOTVKTCTVNSSSLIKKVAVPVK 113
Db 604 --TKPEPSTISDKPAPTPPKGTATTTKEPAPTPPKPAPTPPKGTAPTTLKEPAPTPPK 661
QY 114 KLAQIGTTVTTPKPSVSVAVPTS--VVTTPGKPLNTVTTLKPPSISGASPTSPNE 170
Db 662 KPAPKELAPTTTGGPTSTSDKPAPTPPKPAPTPPKPAPTPPKPAPTPPKPAPTPPKPAPTPPK 716

QY 171 PNKAENSAVQINISPTMLENVKKCNFLMLKLACSSQSPKMGONKAKLYEQULIDA 230
Db 717 PPTSEVSTPTTK-EPTTIH-----KSP----- 739

QY 231 KIEAEFTKRLVELKSSPOPHLVFLKSVALLPNSQSFIOCVQOTSSDMVAT 290
Db 740 ----DEST-----PELSAPTP-----KALENSPKEPGVPT 766

QY 291 CTTVTTSPPVTTTVSSQSEKSIIVSGATAPRVSVQTLNPLAGPVGAKGVYTLHSGV 350
Db 767 TKTPAATKPEMTTAKDKTERDL---RTTPETTT-----AAPKMTKEATTTKTT 815

QY 351 PTAATGCTTAGTGL-----LQTSKPLYSVANVTYVLSL--QPEK 388
Db 816 ESKITATTQVSTTODTTPFKITTKTTLAKVTTTKTITTEIMKPEE 869

RESULT 12

US-10-124-557-58
; Sequence 58, Application US/10124557
; Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cserr, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 58:

SEQUENCE CHARACTERISTICS:

LENGTH: 1049 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 58:

US-10-124-557-58

Query Match 6.9%; Score 156; DB 12; Length 1049;
Best Local Similarity 20.8%; Pred. No. 0.0048;
Matches 86; Conservative 40; Mismatches 164; Indels 124; Gaps 14;

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/10/124,557
  FILING DATE: 16-Apr-2002
  CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/643,502
  FILING DATE: 18-JAN-1991
  APPLICATION NUMBER: US 07/546,114
  FILING DATE: 29-JUN-1990
  APPLICATION NUMBER: US 07/457,196
  FILING DATE: 29-DEC-1989
  APPLICATION NUMBER: US 07/390,901
  FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
  NAME: Gesert, Luann
  REGISTRATION NUMBER: 31,822
  REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (617)876-1170
  TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 14:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 941 amino acids
    TYPE: amino acid
    TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-124-557-14

Query Match      6.9% Score 156; DB 12; Length 941;
Best Local Similarity 20.8%; Pred. No. 0.0042;
Matches 86; Conservative 40; Mismatches 164; Indels 124; Gaps 14;

QY 8 AVPSAPPKVSSGRRLPAPQIVAVKAPNT-----TTIOFPANLQLPPTGYLI 53
DB 482 APTTKREPAPPTPKKPAPTTKETAPPTPKGTAPPTLKKEPAPTKKPAKELAPTT--- 538
QY 54 KNSGSLMLVSPQOVTVAETTSNITSRAVPANPQVAKICVPPNSSOLIKKAVTPVK 113
DB 539 --TKPESTISDKPAPPTPKGTAPPTKKEPAPTTKKEPAPTTKGTAPPTLKKEPAPTTVK 596
QY 114 KLAQIGTVVTVTPKPSVQSAVAPTS---VVTVPGRPLNTVTTLKPSSLGASSTPSNE 170
DB 597 KPAPKELAPTTTKGPTSTSDKPAPTTKETAPPTPKKEPAPT-TPKKP---APTTPEPT 651
QY 171 PNLKENSAAVQINSPMLLENVKKCNFLAMLIKLACSGSSPEMGQNVKRLVEQLDA 230
DB 652 PPTTSEVSTPTTK-EPTTIH-----KSP----- 674
QY 231 KIEAEFTRKLYVELKSSQPHLVPLFKSVVALRQLLPNSQSFIOOCVQOTSSDMVAT 290
DB 675 ----DEST---PELSAETP-----KALENSPKKEGVPT 701
QY 291 CTTVTTSPVTTVTVSSQSEKSIIVSGATAPRTVSVOTLNLAPGVAKAGVTVLHSG 350
DB 702 TKTPATPKREMTTAKDKTTTERDL---RTTPETTT-----AAPKMKETATTEKKT 750
QY 351 PTAATGGTTAGTGL-----LQTSKPLVTSVANTVTIVSL--QPEK 388
DB 751 ESKTATTTQVSTTTQDTTPPKITTLKTTLLAPVTTTKITTTTEIMNKEE 804
```

RESULT 10
US-10-124-557-84

Sequence 84, Application US/10124557
Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.

```
Gesert, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
  ADDRESS: Genetics Institute, Inc.
  STREET: 87 Cambridgepark Drive
  CITY: Cambridge
  STATE: Massachusetts
  COUNTRY: U.S.A.
  ZIP: 02140
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/10/124,557
  FILING DATE: 16-Apr-2002
  CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/643,502
  FILING DATE: 18-JAN-1991
  APPLICATION NUMBER: US 07/546,114
  FILING DATE: 29-JUN-1990
  APPLICATION NUMBER: US 07/457,196
  FILING DATE: 29-DEC-1989
  APPLICATION NUMBER: US 07/390,901
  FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
  NAME: Gesert, Luann
  REGISTRATION NUMBER: 31,822
  REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (617)876-1170
  TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 84:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1022 amino acids
    TYPE: amino acid
    TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-10-124-557-84

Query Match      6.9% Score 156; DB 12; Length 1022;
Best Local Similarity 20.8%; Pred. No. 0.0047;
Matches 86; Conservative 40; Mismatches 164; Indels 124; Gaps 14;

QY 8 AVPSAPPKVSSGRRLPAPQIVAVKAPNT-----TTIOFPANLQLPPTGYLI 53
DB 563 APTTKREPAPPTPKKPAPTTKETAPPTPKGTAPPTLKKEPAPTKKPAKELAPTT--- 619
QY 54 KNSGSLMLVSPQOVTVAETTSNITSRAVPANPQVAKICVPPNSSOLIKKAVTPVK 113
DB 620 --TKPESTISDKPAPPTPKGTAPPTKKEPAPTTKKEPAPTTKGTAPPTLKKEPAPTTVK 677
QY 114 KLAQIGTVVTVTPKPSVQSAVAPTS---VVTVPGRPLNTVTTLKPSSLGASSTPSNE 170
DB 678 KPAPKELAPTTTKGPTSTSDKPAPTTKETAPPTPKKEPAPT-TPKKP---APTTPEPT 732
QY 171 PNLKENSAAVQINSPMLLENVKKCNFLAMLIKLACSGSSPEMGQNVKRLVEQLDA 230
DB 733 PPTTSEVSTPTTK-EPTTIH-----KSP----- 755
QY 231 KIEAEFTRKLYVELKSSQPHLVPLFKSVVALRQLLPNSQSFIOOCVQOTSSDMVAT 290
DB 756 ----DEST---PELSAETP-----KALENSPKKEGVPT 782
QY 291 CTTVTTSPVTTVTVSSQSEKSIIVSGATAPRTVSVOTLNLAPGVAKAGVTVLHSG 350
DB 783 TKTPATPKREMTTAKDKTTTERDL---RTTPETTT-----AAPKMKETATTEKKT 831
QY 351 PTAATGGTTAGTGL-----LQTSKPLVTSVANTVTIVSL--QPEK 388
```

```

      PRIOR APPLICATION NUMBER: US 60/160,587
      PRIOR FILING DATE: 1999-10-20
      NUMBER OF SEQ ID NOS: 440
      SOFTWARE: PatentIn version 3.0
      SEQ ID NO 106
      LENGTH: 1169
      TYPE: PR1
      ORGANISM: Saccharomyces cerevisiae
      US-09-801-368-106

      Query Match      7.1%; Score 159.5; DB 10; Length 1169;
      Best Local Similarity 21.0%; Pred. No. 0.0032;
      Matches 105; Conservative 82; Mismatches 185; Indels 129; Gaps 22;

      QY      4 VTKAAPVAPPKVSSGRLPAPQIVAAKAPNTTIIQF-PANLQLPCTGLIKNSGPIML 62
      DB      689 VTSSSVSTPTTSSSESSASVTIL-----PSTISEFKPSTMK-----TKVSISSPTNL 740
      QY      63 VSPQQTVTRATTTINITSRAV-----PANPQTKICTYPPNSSQILKRVANTPV 112
      DB      741 ITSYDTTSSKDSVTSSVSLSISLIPSSYSASSSEQIFHSSITVSSNGCALTFSSSTKV 800
      QY      113 KK-----LAQIGTVVTTPVK-----PSSVQSAVAVPTSVTVTPGKP 149
      DB      801 SSSSESHSRTPSTSSSGIKSSGVELEISTSPFSEHETSTASTVQSISSQFVTPSSP 860
      QY      150 LNTVTTLKPSLG-----ASTPSNEPLKAENSAAYOINSLPTMLENVKCKKNFLAMLIK 205
      DB      861 ISTVA---PRSTGINSQTESPNSKETWSENSASV-----893
      QY      206 LACGSGOSPENGQWVKLVEQLLDAKIEAE---FTR---KLIVELKSSFPHLVFLK 258
      DB      894 MPSSSATSPKGTG-----KVTSDTSSGFSRDRTVVRMTSETPSTN---EQ 936
      QY      259 KSVVALRLQLNOSQFIQCVQOTSMDVAVTCTTV-----TT-----SPVVTTVSS 307
      DB      937 TTLITVSSCESNCS-----NTVSSAVVSTATTITNGITEYTKMCPISANELTVSK 989
      QY      308 SQS-EKSIIVSGATAPRTVSQVTLNP--LAGPVAKAGVTVLHSGVCTPAATGCTTAGTGL 364
      DB      990 LESEBKTTLLIIVTSCESQVSETSPALIVSTATVNDVYTVVSTWSPQATNKLAVSSD- 1048
      QY      365 LQTSKPLVTSYAMVTYVSLQPEKPVVSGIAVTLSLPAVTGEGTSGAICLPSPKPVVSF 424
      DB      1049 IENSAKSAFVSEAAETKSISSRNNNFVP-TSGTTSIETHT-TTYSNASENSDWNVSASEAV 1106
      QY      425 CWDHCKRPVI-----GTP 437
      DB      1107 SSKSVTNPLVLSVSOQPRGTP 1127

      RESULT 9
      US-10-124-557-14
      : Sequence 14, Application US/10124557
      : Patent No. US20020137894A1
      : GENERAL INFORMATION:
      : APPLICANT: Turner, Katherine
      : Clark, Stephen C.
      : Jacobs, Kenneth
      : Hewick, Rodney M.
      : Gesner, Thomas G.
      : TITLE OF INVENTION: Megakaryocyte Stimulating Factors
      : NUMBER OF SEQUENCES: 143
      : CORRESPONDENCE ADDRESS:
      : ADDRESSER: Genetics Institute, Inc.
      : STREET: 87 CambridgePark Drive
      : CITY: Cambridge
      : STATE: Massachusetts
      : COUNTRY: U.S.A.
      : ZIP: 02140
      : COMPUTER READABLE FORM:
      : MEDIUM TYPE: Floppy disk
      : COMPUTER: IBM PC compatible
  
```


US-09-893-519A-14

Query Match 22.8%; Score 513; DB 9; Length 1023;
Best Local Similarity 33.4%; Pred. No. 1.3e-27;
Matches 154; Conservative 62; Mismatches 127; Indels 118; Gaps 19;

QY 13 PRVSSG-----PRLPAQIVAKAPNTTIOFPANIQLPGLVILKSNGLM 61
DB 414 PRATSGIRATLPVLABRLPQ-----PQNPNIQ---NFQLPGLVLSNGQLL 464
QY 62 LVSPOQTVR-----AETSNTSRPAVPANDQTKICTVPRSSQLIKKAVATYVKL 115
DB 465 MI-PQALAKOMQAQAQOTTMAPRPTSPAPVOISTVAPGPTIAR-QVTP----- 518
QY 116 AIGITVTVTPKPSVQSAVPTSVYTPGKPLNTYT--TLKPSLSGSS-----TPS 168
DB 519 -----TIIIOV---SQAQITVQPSATLQSPGQPOLVIGGAQTSLSGATATVQGTQ 571
QY 169 NE-PNKAENSAVQINISPTMLENVKCKNPLAMLIKACSGSGSPERGQNVKLEVDL 227
DB 572 RVPFGATTSSAATE-----TMEVYKCKNFKLSTLIKIASGKQSTETAAVKELOVNL 625
QY 228 LBAKTEAEFTKRLYEKSSPOPHVPLKKSVALRQLRPSQSFIOQCVO-----TS 283
DB 626 LQKTEAEFTSKRLKELSSPOPHVPLKKSVALRQLRTPSAFIOQSOQPPPPS 685
QY 284 SDMVATCTTWTTSVPTTTSVSSQSEKSIYSAGTAPRVSQTLNPLAGPVAKAGV 343
DB 686 Q-----ATTALNAVLSSSVQRTAGKTAATVTSALQPPVLS----- 722
QY 344 VTLHSVGPAAATGCTAGTGLQTSKPLVTSVANTVTVSLOPEKPVSGTAVTSLPAY 403
DB 723 -----TOPTVGVGKQSQPTPLVIO-----QPPKP-----GALIRPQV 756
QY 404 TGETSGAICLPVSKPVVSVFCMDHICKPVIGTPOIKLAQ 444
DB 757 TLTOT-----PMVALRQPH-NRIMLTTPQOVNLS 785

RESULT 2

US-09-801-368-108
Sequence 108 Application US/09801368
Patent No. US20020128250A1
GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Call, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Milne, Todd
APPLICANT: Maxon, Mary
APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amir
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801,368
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US 60/160,587
PRIOR FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 108
LENGTH: 1367
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-801-368-108

Query Match 8.9%; Score 200.5; DB 10; Length 1367;
Best Local Similarity 22.2%; Pred. No. 6e-06;
Matches 113; Conservative 69; Mismatches 225; Indels 101; Gaps 17;

QY 2 TLVTKAVPAPKPVSSGPRLPAP-----QIVANKAPNTTIOFPANIQLPGLVILK 54
DB 618 TRESSEAPVYSSSTTESSAPVPPSSSTTESSAPVPPSSSTTESSAPVPPSSSTTE 677
QY 55 SNGPMLVSPQOT---VRAETSNITSRPAVPANDQTKICTVNNSSQLIKKAV- 109
DB 678 SSSAPVSSSTTESSAPVSSSTTESSAPVPPSSSTTESSAPVPPSSSTTESSAPV 737
QY 110 -PPVKIAQIGTIVTT-----VPRSSVQO-----VAVPTSVTVTPGKPLNT- 152
DB 738 PRSSSTTESSAPVTSSTTESSAPVPPSSSTTESSAPVPPSSSTTESSAPVPPSS 797
QY 153 -----YTLKPSLSGASSTPSNEPNKAENSAVQINISPTMLENVKCKNFIA 201
DB 798 SSSSTESSVAPVPTPSSSNITSAPSSTPSSSTESSVPPV---PTPSSSTESS- 851
QY 202 MLIKLACSGSOS-----PEMGQNVKLEQLDAKIEAEFTKRLYEKSSPQPHLY 254
DB 852 --APVSSSTESSVAVPPPPSSSNITSAPSSTPSSSTTESSTGTTVTPSSSK----- 904
QY 255 PFLKKSVALRQLRPSQSFIOQCVOOTSDVATCTTWTTSV---VTTTVSSSQSE 311
DB 905 -----YPSQOT--ETSVSTETTTIVPKTTTSVTPPTTTTITTVCGSTGN 949
QY 312 KSITVSGATAPRV---SVQTLNPLAGPVAKAGVTLHSVGPAAATGCTAGG--LLOT 367
DB 950 SAGETTSAGSPKTVTTVTPPTTTTSVTSSTTTTITTVCGSTGNSAGETTSAGSPKTTT 1009
QY 368 SKPLTVSVANT-----VTVLSLOPEKPVSGTAVTSLPAY--FGTSGAIC----- 414
DB 1010 TVPCSTSPSTSESTTSTSTPTTVTVSTTVTTTSTSTKGGELITTVTKNIPPT 1069
QY 415 -----LPSVYPVSVFCMDHICKPVIGT 436
DB 1070 LTTIAPTPSVTVTVNFTPTTITTVCGST 1097

RESULT 3

US-10-025-380-1068
Sequence 1068 Application US/10025380
Publication No. US20020182191A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tongtong
APPLICANT: Jiang, Yugu
APPLICANT: Smith, Carole L.
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Skelley, Yasir A. W.
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick Thomas S.
APPLICANT: Carter, Darrick
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
FILE REFERENCE: 210121.471C14
CURRENT APPLICATION NUMBER: US/10/025,380
CURRENT FILING DATE: 2001-12-19
NUMBER OF SEQ ID NOS: 1129
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1068
LENGTH: 5179
TYPE: PRT
ORGANISM: Homo sapiens

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:56:34 ; Search time 8.9045 Seconds
(without alignments)
1296.883 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_452
Perfect score: 2249
Sequence: 1 GRLVTKVAPVAPKSSCP.....VICTPVQIKLQEPVLISQ 452

Scoring table: BLOSUM62

Gap: 10.0 , Gapext 0.5

Searched: 140259 seqs, 2554876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB pep: *
2: /cgn2_6/ptodata/2/pubppa/PCF_NEW_PUB pep: *
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB pep: *
4: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB pep: *
5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB pep: *
6: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB pep: *
7: /cgn2_6/ptodata/2/pubppa/PCFUS_PUBCOMB pep: *
8: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB pep: *
9: /cgn2_6/ptodata/2/pubppa/US09_NEW_PUB pep: *
10: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB pep: *
11: /cgn2_6/ptodata/2/pubppa/US10_NEW_PUB pep: *
12: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB pep: *
13: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB pep: *
14: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	513	22.8	1023	9 US-09-893-519A-14	Sequence 14, Appl
2	200.5	8.9	1367	10 US-09-801-368-108	Sequence 108, Appl
3	195.5	8.7	5179	9 US-10-025-380-1068	Sequence 1068, Ap
4	195.5	8.7	5179	10 US-09-922-217-1068	Sequence 1068, Ap
5	195.5	8.7	5179	10 US-09-833-263-1068	Sequence 1068, Ap
6	172	7.6	1322	10 US-09-801-368-114	Sequence 114, Appl
7	160	7.1	2665	10 US-09-864-761-34248	Sequence 34248, A
8	159.5	7.1	1169	10 US-09-801-368-106	Sequence 106, Appl
9	156	6.9	941	12 US-10-124-557-14	Sequence 14, Appl
10	156	6.9	1022	12 US-10-124-557-84	Sequence 84, Appl
11	156	6.9	1038	12 US-10-124-557-74	Sequence 74, Appl
12	156	6.9	1049	12 US-10-124-557-58	Sequence 58, Appl
13	156	6.9	1075	10 US-09-801-368-110	Sequence 110, Appl
14	156	6.9	1140	12 US-10-124-557-104	Sequence 104, Appl
15	156	6.9	1270	12 US-10-124-557-44	Sequence 44, Appl
16	156	6.9	1311	12 US-10-124-557-42	Sequence 42, Appl
17	156	6.9	1313	12 US-10-124-557-142	Sequence 142, Appl
18	156	6.9	1314	12 US-10-124-557-50	Sequence 50, Appl
19	156	6.9	1320	12 US-10-124-557-46	Sequence 46, Appl

20	156	6.9	1320	12 US-10-124-557-60	Sequence 60, Appl
21	156	6.9	1354	12 US-10-124-557-48	Sequence 48, Appl
22	156	6.9	1361	12 US-10-124-557-40	Sequence 40, Appl
23	156	6.9	1363	12 US-10-124-557-52	Sequence 52, Appl
24	156	6.9	1404	12 US-10-124-557-2	Sequence 2, Appl1
25	156	6.9	1404	12 US-10-124-557-62	Sequence 62, Appl
26	153	6.8	1601	10 US-09-862-027-40	Sequence 40, Appl
27	152.5	6.8	386	10 US-09-864-761-35720	Sequence 35720, A
28	150	6.7	1056	9 US-10-161-510-10	Sequence 10, Appl
29	150	6.7	1537	10 US-09-801-368-104	Sequence 104, Appl
30	145	6.4	688	10 US-09-864-761-36047	Sequence 36047, A
31	142.5	6.3	2586	10 US-09-905-129-11	Sequence 11, Appl
32	142.5	6.3	2586	10 US-09-905-129-14	Sequence 14, Appl
33	142.5	6.3	2586	10 US-09-905-129-11	Sequence 11, Appl
34	142.5	6.3	2586	10 US-09-905-129-11	Sequence 11, Appl
35	142.5	6.3	2587	10 US-09-905-129-16	Sequence 16, Appl
36	142.5	6.3	2587	10 US-09-905-129-16	Sequence 16, Appl
37	142.5	6.3	2589	10 US-09-905-129-24	Sequence 24, Appl
38	141	6.3	2597	10 US-09-905-129-2	Sequence 2, Appl1
39	141	6.3	2597	10 US-09-905-129-10	Sequence 10, Appl
40	141	6.3	2597	10 US-09-905-129-13	Sequence 13, Appl
41	141	6.3	2597	10 US-09-905-129-13	Sequence 13, Appl
42	141	6.3	2597	10 US-09-905-129-13	Sequence 13, Appl
43	141	6.3	2597	10 US-09-905-129-13	Sequence 13, Appl
44	139	6.2	2005	10 US-09-735-3678-3	Sequence 3, Appl1
45	139	6.2	2063	10 US-09-735-3678-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1

US-09-893-519A-14
Sequence 14, Application US/09893519A
Publication No. US20030027243A1

GENERAL INFORMATION:

APPLICANT: ANADYS PHARMACEUTICALS, INC.

APPLICANT: THOMPSON, Craig

APPLICANT: MOORE, Jeffrey

APPLICANT: BUTRMAN, Ed T.

APPLICANT: BRADLEY, John

APPLICANT: DESILVA, Thamara

APPLICANT: HARRIS, Sandra

APPLICANT: KOMARNITSKY, Svetlana

APPLICANT: MENDILLO, Marc

APPLICANT: MOORE, Daniel

APPLICANT: MCCOY, Melissa

APPLICANT: SANDERSON, Karen

APPLICANT: HAO, Tariq

APPLICANT: ZHU, Shuhao

APPLICANT: LONG, Fan

APPLICANT: DAVIDOV, Eugene

TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE

FILE REFERENCE: 0342/1G548-US2

CURRENT APPLICATION NUMBER: US/09/893,519A

CURRENT FILING DATE: 2001-06-28

PRIOR APPLICATION NUMBER: US 60/215,164

PRIOR FILING DATE: 2000-06-29

PRIOR APPLICATION NUMBER: US 60/224,457

PRIOR FILING DATE: 2000-08-10

NUMBER OF SEQ ID NOS: 146

SOFTWARE: PatentIn version 3.1

SEQ ID NO 14

LENGTH: 1023

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

OTHER INFORMATION: Corresponds to SEQ ID NO: 87

PUBLICATION INFORMATION:

DATABASE ACCESSION NUMBER: Human Genbank/CAA72189

DATABASE ENTRY DATE: 1997-06-25

RELEVANT RESIDUES: (1)...(1023)

This Page Blank (uspto)

Db 255 SLANO-OLPPACARQLSKRLTLQOF--GNDISPEIGERYTLVLGVNSTLTIE 311
QY 236 EFTKRLVELSSPOPHLYPFLKSVVAL-ROLLENSQSFIO---QCVQOITSSDMVIATC 291
Db 312 EFHSLKEATNEPLRPFVLPFLKANLPLORELLHCARLAKONPAQYLAHOEQLLDAST 371
QY 292 TTTVTTSFV 301
Db 372 TSPVDSSELL 381

RESULT 15
US-08-325-267A-4
; Sequence 4, Application US/08325267A
; Patent No. 5585271
; GENERAL INFORMATION:
; APPLICANT: WATARI, JUNJI
; APPLICANT: TAKATA, YOSHIHIRO
; APPLICANT: OGAWA, MASAHIRO
; APPLICANT: PENTTILA, MERJA
; APPLICANT: ONNELA, MAIJA-LEENA
; APPLICANT: KERANEN, SIKKA
; TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST
; TITLE OF INVENTION: CONTAINING THEM
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325.267A
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP PCT/JP94/00290
; FILING DATE: 24-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 38871/1993
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2589-023-0XPCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SRO ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 862 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-325-267A-4

Query Match 7.3%; Score 163.5; DB 1; Length 862;
Best local similarity 21.1%; Pred. No. 1.4e-05;
Matches 102; Conservative 67; Mismatches 156; Indels 159; Gaps 21;

QY 11 SAPKXSSGRLLPAPQIVAAVAPNTTIQF---PANQLPPTGVLIKNSGRL----- 60
Db 199 SLPEITGTYMAYGYYPKAVVSNVSWGTLPISVTLPGTGVYSDDFEGYVVSFDDDL 258
QY 61 -----MLVSPQQTVTTRAET-----TSNITRPAVPAVPAVQIVKICTVPS 99

Db 259 SOSNCTVPDPNSYAVSTTTTTTEPWTGFTSTSTEMTIVTGNGVPTD-ETVIVIRPTT 317
QY 100 SSQILKVAVTPVKRLAQIGTIVTTVPKSSVQSAVFTSVTVTPGKPLN-TVTTIK- 157
Db 318 AS-----TITTEPWTGFTSTSTEMTIVTGNGVPTD-ETVIVIRPTT 359
QY 158 PSSLGASSPNSNP-----NIKENSAAVQINLSPTMLENVKKKNPLAMIKIACGS 211
Db 360 PTSEGLVTT-TTEPWTGFTSTSTEMTIVTGNGVPTD-ETVIVIRPTT-PTNAISSSSSS 416
QY 212 QSPENGQVKKLYEQLLDKAKIEEFTKLYELKSSPOPHLYPFLKSVVALROLLENS 271
Db 417 SSGQITSSI-----TSSRPITTPPPS-----NG 440
QY 272 QSFIOQCVQOITSSDMVIATCTTV-TTSPVTTTVSS-----SQSEKSIIV---- 316
Db 441 TSVI-----SSSVISSVTSLSFTSPVSISSVTSSTSTSTSTSTSTSTSTSTSTST 493
QY 317 --SGATAPRTVSQTLNPLAGPYGAKAGVYTLHSGPPTAAGCTTAGTGLQTSKPLVTS 374
Db 494 STSGSSESTSS-----AGSVSSSSSFSSSSSKSPYSS-----SSLPVTS 535
QY 375 VANTVTVSLQPEKRPVSGTAVTSLPNAVTFGETSGAICLPSVKNPVVSCMDHICKPVI 434
Db 536 ATTSQETAS-----SLPATTTTKTSEQTLV-----TVTSCSHVCTESI 575
QY 435 GTPV 438
Db 576 SPAT 579

Search completed: February 16, 2003, 22:02:50
Job time : 20.6146 secs

QY 397 TSLPAP 403
 Db 677 TTPLPDI 683

RESULT 13

US-08-928-361B-6

; Sequence 6, Application US/08928361B
 ; Patent No. 6071518

; GENERAL INFORMATION:

; APPLICANT: Petersen, Carolyn

; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
 ; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS

; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PETERS, VERNY, JONES & BIKSA

; STREET: 385 Sherman Avenue, Suite 6

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94306-1840

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/928,361B

; FILING DATE: 12-SEP-1997

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/026,062

; FILING DATE: 13-SEP-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: VERNY, Hana

; REGISTRATION NUMBER: 30,518

; REFERENCE/DOCKET NUMBER: 480,76-1(HV)

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-324-1677

; TELEFAX: 650-324-1678

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1721 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-928-361B-6

Query Match 7.6%; Score 171.5; DB 3; Length 1721;
 Best Local Similarity 21.5%; Pred. No. 8.3e-06;

Matches 92; Conservative 50; Mismatches 198; Indels 87; Gaps 15;

QY 35 TTTTTPPANLQPLPGTGLIKNSGPMVLVSPQOTYRAETTSNITSRPAVPANPQTVKIC 94

Db 286 TTTTTPPANLQPLPGTGLIKNSGPMVLVSPQOTYRAETTSNITSRPAVPANPQTVKIC 94

QY 95 TYPNSSOLIKKVAATPVAKKLAIGTIVTVTPKPSVSVAVPVSIVVTTEKPLANTV 154

Db 346 TTTTTPPANLQPLPGTGLIKNSGPMVLVSPQOTYRAETTSNITSRPAVPANPQTVKIC 94

QY 155 TLPSLSGASTPS-----NEPNLKAENSA-----VOINLSPMLN----- 192

Db 399 TTAATTTTSETESVIAKPEKMCWLENGCEAKGATVAVVIGKDGRIENGMAFTMLPND 458

QY 193 -----VKCKNPLAMLIKILACSGSGSPENGQNVKLVLEQLDAKIE--AEETFKLY 242

Db 459 THVRFKVKDVGNITISVCRKAGKLEP-----DRSLDETPVAGHNSCSII 508

QY 243 VLLKSSPOHVLPLKKSVALROLPLNOSF-----IQCV-----QOTSSMVA 289

Db 509 VGVSGDGKIHVSYPGSKDVLISAPIQSELEFNEVYCDCTAKYGAHSGQTSADFEVT 568
 QY 290 TC---TTTWTSPVYTTTSSOSEKSIIVSGATAPRTVSOTLNPAGVGAAGVTL 346
 Db 569 TTAFTTTTGGAGQPTTTTGSPPK---TTTTTATTTTTTLN-----ITTT 616
 QY 347 HSYPTMATGTTAGTGLLQTS---KPLVTSVAN--TTVTSLOPEKPVSG---TAV 396
 Db 617 TTQKPTTTTTRKVPKRPATTTTLKPLVTTTTRKATTTTTTTTTRKDMTIT 676

QY 397 TSLPAP 403
 Db 677 TTPLPDI 683

RESULT 14

US-08-244-189-2

; Sequence 2, Application US/08244189

; Patent No. 5580727

; GENERAL INFORMATION:

; APPLICANT: Ohki, Misao

; APPLICANT: Kikuchi, Kimiko

; APPLICANT: Miyoshi, Hiroyuki

; TITLE OF INVENTION: No. 5580727el DNAs, Polypeptides Encoded Thereby

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Birch, Stewart, Kolasch & Birch

; STREET: 8110 Gatehouse Road, Suite 500 East

; CITY: Falls Church

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22042

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/244,189

; FILING DATE: 15-AUG-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Murphy Jr., Gerald M.

; REGISTRATION NUMBER: 28,977

; REFERENCE/DOCKET NUMBER: 760-183P

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 205-8000

; TELEFAX: (703) 205-8050

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 752 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-244-189-2

Query Match 7.3%; Score 164.5; DB 1; Length 752;
 Best Local Similarity 28.8%; Pred. No. 9.7e-06;

Matches 72; Conservative 41; Mismatches 108; Indels 29; Gaps 11;

QY 61 MLVSPQOTYRAETTSNITSRPAVPANPQTVKICVTPNSSOLIKKVAATPVAKKLAIGT 120

Db 152 VFTNPQVATYHRAIKITVDGPREPN-RTERKHSITPDSPVDVKTQSRLLP----- 201

QY 121 TTTTTPKPSVSVAVPTSVTVTPGKPLNTVTLKPSL-GASTP---SNEPLKAE 176

Db 202 ---PTMPPTTQG-APRISSTPT---TLTNGTSHSPALNAGAPBPFGFSGSSSS 254

QY 177 NSANVOINISPTM-LENVKKCNFLAMLIKILACSGSGSPENGQNVKLVLEQLDAKIEA 235

```

Oy 156 LKPS-----SLGSSSPSENEPNLKKENSAAYQINISPTMLENVKKCKKNLAMI 204
Db 1120 LSPDQARLPSEGVVSIQGLASLAQRPVANNAGSGPLPTFQIQGN----- 1163
Oy 205 KLASGSGQ-----SPEGNQVKKLIVREOLDLDAKLEAEFEETKLYLEUKSSQPHLPEFL 257
Db 1164 KTLTGAGVQROLAVGQPRQLQMPPIVMVNTGVYKLVNQAPRD-----GLTFVPPILANP 1218
Oy 258 KKSVALNQLLENPSQSFIOQCVQOTSSDMVIATCTVTTSFVVTY-----VSSSQ 309
Db 1219 RPPSSGLPAVLN-----PRPLTFTGRLPTPLTGARAPMPPTPLVRPLKLTVHSPS 1269
Oy 310 SEKSLIVGATPRTVS-----VQNLNPLAGVGNKACY-----V 344
Db 1270 PEVSASARCAAPLITISPLHVPSSLPGPASSPMDIPNSSPLASVSSSTVSVLSSLP 1328
Oy 345 TLHSVGPATAGTGT-----AGTGGLD-QTKPLVTSVANTVTVTSLQPEKPVVS 392
Db 1329 SVPTTLNPAASAPLPIIPISAPLTVSASGPAALLTSTPTPLAPVPVPAAPGPPSIQPSGASPS 1388
Oy 393 GTAAVTLST---PAVTFGETSGAICL-----PSVKPVSECMWDHICKPVYIGTFVQIKL 442
Db 1389 ASALTLGLATAPSLTSSQTPGHPHLLLAFTSHVPLGNTSVA-----PACSPVL-VPAS-AL 1442
Oy 443 AQQPGVLSQP 452
Db 1443 ASPPPSADNP 1452

```

```

RESULT 11
US-09-165-239A-4
: Sequence 4, Application US/09165239A
: Patent No. 6344554
: GENERAL INFORMATION:
: APPLICANT: JOHNSON, ALEXANDER
: APPLICANT: BRAUN, BURKHARD R
: TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES FROM CANDIDA
: TITLE OF INVENTION: ALBICANS ENCODING POLYPEPTIDES ASSOCIATED WITH FILAMENTOUS
: TITLE OF INVENTION: GROWTH
: FILE REFERENCE: 220022000700
: CURRENT APPLICATION NUMBER: US/09/165,239A
: CURRENT FILING DATE: 1998-10-01
: PRIOR APPLICATION NUMBER: 60/068,065
: PRIOR FILING DATE: 1997-12-18
: NUMBER OF SEQ ID NOS: 13
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 4
: LENGTH: 750
: TYPE: prt
: ORGANISM: Candida albicans
US-09-165-239A-4

```

Query Match	7.6%;	Score 172;	DB 4;	Length 750;
Best Local Similarity	23.6%;	Pred. No. 2.2e-06;		
Matches 106;	Conservative 53;	Mismatches 176;	Indels 114;	Gaps 20

```

OY 4 VTKAVPAPPKVSSGRLPAPOLAVAKAPMTTITIOEPANIQLPBGTVILKNSGMLLV 63
Db 347 VPSTTPSSAPETTPSSAPSSVPSSAPESAPETTPSSAPSSV-----ESSAP--- 395
OY 64 SPOQTVAETTSNTSRKAPAPDQVTKICTYVNSSOLIKKAVATPVKKAQIGTVV 123
Db 396 ---ELETETPTLAHTTTTA-----QTTTVITVISCNNMCKEVI-----TGVVV 440
OY 124 T-----TVKPSVSVAVPVISVTVTPGKPLNTVTTLKPSLGLASSTP-S 168
Db 441 TSEDITVYTFCLPETTPPVPSVDSTSV-TSAPETTP-----ESTAPESAPSSAPES 493
OY 169 NEPNLKAENSAAYOINISPTLAEVWKCKKNLAMLKLAGSGSSPEMGQNVKKLYBQL 228
Db 494 SAPYETETPGVSYVTEBSKITVITITSCN-----NACSES-----KYTTGVV 536
OY 229 DAKIEAEETFRKLVEL-----KSSQPHLVPLFLKKSVALROLLPNSQSFIO 276

```

```

Db      537  --VTSSEYVTTTCPLLETPATESAPESAPPTESAPTESAPTESAPVAPESAPPETET--- 591
QY      277  QCVOQTSDDMIATCTTCTVTTSPTVTTTSSSOSEKSLTSGATAPRTVSQTLNPLAGP 336
           :::::  |:::  |:::  |:::  |:::  |:::  |:::  |:::  |:::  |:::
Db      592  ----APATESAPAT-----ESSPVAPEGSESPVAPESAPPTESAPTES----SPVA-- 633
QY      337  VGAKAGVYTLHSVGPNTATG---TTAGTGLQTSKPLVTS--VAANYTVTVSLQPEKPV 391
           |:::  |:::  |:::  |:::  |:::  |:::  |:::  |:::  |:::
Db      637  ----PGTET-----TPATPPEASTPVAAPVAPESAPVAVESSPVAQGVETTPVAAPVAP-- 684
QY      392  SGTAVTSLPAPVTEGETGAICLPSVAP 420
           |:::  |:::  |:::  |:::  |:::  |:::  |:::  |:::  |:::
Db      685  STAKTSALVSTEGTETPTTLESYPAIDP 713

```

```

RESULT 12
US-08-700-651-5
: Sequence 5, Application US/08700651B
: Patent No. 6015882
: GENERAL INFORMATION:
: APPLICANT: PETERSEN, CAROLYN
: APPLICANT: LEECH, JAMES
: APPLICANT: NELSON, RICHARD, C.
: APPLICANT: GUY, JIRI
: TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
: TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
: TITLE OF INVENTION: INFECTIONS
: FILE REFERENCE: 480.19-4(HV)
: CURRENT APPLICATION NUMBER: US/08/700,651B
: CURRENT FILING DATE: 1997-08-14
: EARLIER APPLICATION NUMBER: 08/415,751
: EARLIER FILING DATE: 1995-04-03
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 5
: LENGTH: 1721
: TYPE: PR1
: ORGANISM: Cryptosporidium parvum
US-08-700-651-5

```

Query Match	7.68	Score 171.5	DB 3	Length 1721
Best Local Similarity	21.5%	Pred. No. 8.3e-06		
Matches 92	Conservative 50	Mismatches 198	Indels 87	Gaps 15

QY	35	TTTTIOFPANLQLPGGVILIKNSGSPMLVSPQGVTRAETTSNITSRAVPANPQIVKIC	94
Db	286	TT	345
QY	95	TVPNSSQLIKNAVTPVKRLAQIGTVVTVVPRPSSVQSAVETSVYVYTPGRLTVY	154
Db	346	TT	399
QY	155	TLKSSSLGASSTPS-----NEPNLKENSAA-----VOINLSPMLEN--	199
Db	399	TTATTTTTTSTESVYIKPKDECMCEKNECCAKATVGVYIGKRGRIENGMAFTMIENDD	455
QY	193	-----VKKCKNEFLAMLILACSGSOSPMSGONKKVLEQLDAKIE--AEETRKLY	242
Db	459	THVRFRFKVKNQVGNVTSVRCRKGAGKLEFP-----DRSIDETIPPVAGHNSCII	508
QY	243	VELKSSPPHLYVPELFKSVYALRQLLEPNSQSF-----IQGCY-----QQTSSDVYA	289
Db	509	VGVSGDGKIHSPYSGSKDVSLSNAPIDQSELENEVCDTCAKAYGAISHGVQTSADCVTT	568
QY	290	TC-----TTVTVPSPVTTTTVSSSOSSEKSLIVSGATAPRTSVQTLNPLAGPVAKAGAVTL	348
Db	569	TTAKPTTTTTTCAPOQPTTTTTGSSKRP--TTTTTKATTTTTTLNP-----IITF	616
QY	347	HSVGPFLAATGGTACTGLQLOTS--KPLVTSVAN--TVTVVSLQRPKPVVSG--TAV	386
Db	617	TTQKPTTTTTTTKVVGKPPPIATTTTTLTKTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	678

PCT-US93-11721-5

Query Match 9.1%; Score 204; DB 5; Length 2035;

Best Local Similarity 23.5%; Pred. No. 1.7e-08; Matches 119; Conservative 66; Mismatches 188; Indels 134; Gaps 24;

```
QY 10 VSAPPKVSSGPRLPAPQIVAKAPNTTTIOFPANLQLPFGTVLTKNSGRLMVSPOQTV 69
DB 537 IGSPPMSGMAALAAAAATOKIPSSA---PVLVSVPAGTITVKT-----MAVTEGTTT 588
QY 70 TSEETSNITSRPANPQATKICVPRSSSSQKLVKAVTPVKKLAQIGTV----- 122
DB 589 LPA---TVKVASPSVMSNAT-----RMLKTAAL-----AQVETSASATNTS 628
QY 123 ---VTVPRSSVSAVPTSVTVTPGKPLNTVTLK-PSSL-GASTPSNEPNL----- 173
DB 629 TSPITVHKSGTV-TVAQQAQVTVTVGGVTKITILVKRPIVSGSALISNIGKMSV 687
QY 174 --KAENSAAVQINLSPTMLENVKCKKNFL--AMLKILACSGSOSP-----EMCONVK 221
DB 688 QTRPVQTSVNTGQASGPTQIOTKGPAGTILKLVTSADGKPTTITTTQASGATK 747
QY 222 KIVEOLDLAKIEEFTRLVLELKSSPQPHLVPLFKKSVVALRQLPMSQSFIOQCVOQ 281
DB 748 PTLTIGI-----SSVSPSTT--KPGTTTITKIPMSAITTQAGANG 785
QY 282 TSSDMVIATCTTVT-----SPVVT-----TVSSSQSEKSIIVSGA-----TA 321
DB 786 VSSSGIKRPITITTKVMTSGTAPAKIITAVPKIATHGQOGTVQVVLKAPGPGPTI 845
QY 322 PPT-----VSQTLNPLAGPVAK--AGVTLHSVGPFAVGTGTAQTLGQTS 368
DB 846 LRTVPMGVRVLTPTVVSIVKPAVTLTVKGTGVTTLCTVGTGST--SLAGAGSHS 903
QY 369 KPLVTSVA--NTVTYSIQPEKPVVSTAVTLSLPAVTEGSGAICLPSVKEPVVSPFC 426
DB 904 ASLAPITTLTGITATLSSQ---VINPTAVISAQTTTLAAGGLTPTTITMQPV----- 954
QY 427 DHICKRVIGTPVOIKL-AQPGVLSOP 452
DB 955 -----SQPTQVTLITAPSGVEAOP 973
```

RESULT 9
US-09-579-181-2

; Sequence 2, Application US/09579181

; Patent No. 6365372

; GENERAL INFORMATION:

; APPLICANT: Chivria, John

; APPLICANT: Yaciuk, Peter

; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)

; FILE REFERENCE: 16153-4247

; CURRENT APPLICATION NUMBER: US/09/579,181

; PRIOR APPLICATION NUMBER: 60/136,620

; PRIOR FILING DATE: 1999-05-27

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 2

; LENGTH: 2972

; TYPE: PRT

; ORGANISM: Human

US-09-579-181-2

Query Match 7.8%; Score 176.5; DB 4; Length 2972;

Best Local Similarity 22.4%; Pred. No. 6.9e-06;

Matches 123; Conservative 56; Mismatches 196; Indels 175; Gaps 26;

```
QY 9 PVSAPPKVVSSGPRL-----PAPQIVAKAPNTTTIOFPANLQL-----P 47
DB 826 PYRPPP-----GPELSAQPTPGVPQV-----PASLMVSASAPGAPPLIPASRP 869
QY 48 PCTVL---IKNSNG--PLMLVSPQOTVTRAETTSNITSRPV-----PANQTVKICTV 96
```

```
DB 870 PGVLLPLPQPNSSGLPQVLPSPDLGVLSC-----TSRPPTTSLIKPFPAPVRLSPA 922
QY 97 -PNSSQLIKKVAIVPVKKLAQIGTVTVTPRPSVQSAVPTSVTVTPGKPLNTVTT 155
DB 923 PPGSSSLIKPLVPLPGYFPAPAAATTTSTTATATTAVPAT-----PADORLI 973
QY 156 LKPS-----SLGASTPSNEPNLKAENSAVQINLSPTMLENVKCKKNFLAMLI 204
DB 974 LSPDMQARLPSSGVRVSIQGLASLAQRPVANAGSKRLFTQIQGN----- 1017
QY 205 KLAGSGSQ-----SPENQONVKLVEOLDLAKIEEFTRLKLVLELKSSPQPHLVPL 257
DB 1018 KLTLTGAQVROLAVGQPRPLQMPPTVMVNTGVYKIVVROAPRD-----GLTVPPLAPAP 1072
QY 258 KKSVALRQLPSPSSQFIQCCVQGTSSDMVIANCTTVTTSPTVTT-----VSSQ 309
DB 1073 RPPSSGLPAVLN-----PPTTLPGRLPPTLTGTAAPAPPTTLVRLIKLVHSPS 1123
QY 310 SEKSIIVSGATAPRTVS-----VQTLNPLAGPVAKAGV-----V 344
DB 1124 PEVSASAPGA-APLTISSPLHVPSSLGPASSPMPIPNSSPLASPVSSIVSVLSSSLP 1182
QY 345 TLHSVGPPLAATGTT-----AGTGLL-QTSKPLVTSVANTVTYSIQPEKPVVS 392
DB 1183 SVPTTLPAAPASAPLTIPISAPLTVSASGPAALLTSVTPPLAPVPAPAGPSSLQPSGASPS 1242
QY 393 GTAVTLSTL---PAVTFGSGAAILC-----PSVKPVVSPFMDHICKRVIGTPVOIKL 442
DB 1243 ASALVTGLTAPBSLSSSQTPGHPDLAPITSSHAVGLNSTVA-----PACSPVL-VPAS-AL 1296
QY 443 AOPGPVLSOP 452
DB 1297 ASFPSPAPNP 1306
```

RESULT 10
US-09-579-181-1

; Sequence 1, Application US/09579181

; Patent No. 6365372

; GENERAL INFORMATION:

; APPLICANT: Chivria, John

; APPLICANT: Yaciuk, Peter

; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)

; FILE REFERENCE: 16153-4247

; CURRENT APPLICATION NUMBER: US/09/579,181

; PRIOR APPLICATION NUMBER: 60/136,620

; PRIOR FILING DATE: 1999-05-27

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 1

; LENGTH: 3118

; TYPE: PRT

; ORGANISM: Human

US-09-579-181-1

Query Match 7.8%; Score 176.5; DB 4; Length 3118;

Best Local Similarity 22.4%; Pred. No. 7.5e-06; Matches 123; Conservative 56; Mismatches 196; Indels 175; Gaps 26;

```
QY 9 PVSAPPKVVSSGPRL-----PAPQIVAKAPNTTTIOFPANLQL-----P 47
DB 972 PYRPPP-----GPELSAQPTPGVPQV-----PASLMVSASAPGAPPLIPASRP 1015
QY 48 PCTVL---IKNSNG--PLMLVSPQOTVTRAETTSNITSRPV-----PANQTVKICTV 96
DB 1016 PGVLLPLPQPNSSGLPQVLPSPDLGVLSC-----TSRPPTTSLIKPFPAPVRLSPA 1068
QY 97 -PNSSQLIKKVAIVPVKKLAQIGTVTVTPRPSVQSAVPTSVTVTPGKPLNTVTT 155
DB 1069 PPGSSSLIKPLVPLPGYFPAPAAATTTSTTATATTAVPAT-----PADORLI 1119
```

Db 748 PTLIGT-----SSVSPST---KPGTTITIKTIMSAITIQAGATG 785
QY 282 TSSDMVATCTTIVTT-----SPVVT-----TTVSSSQSEKSIIVSGA-----TA 321
Db 786 VTSSPGIKSPITITITTKMTSGTGAPAKIITAVPKIATIGHQOGVTVYVLGABQGPBTI 845
QY 322 PRT-----VSQTLNPLAGPVGAK--AGVTLHSVGPPTAAGTTAGTGLQTS 368
Db 846 LRTVPMGCVRLVPTVSAVAPVATTVLVKGTGTGTVTGTVST--SLAGAGHSTS 903
QY 369 KPIVTSVA--NTVTVTSIQPEKPVVSGTAVTSLPAVTFGETSGAALCLPSVKPVVSRGW 426
Db 904 ASLATPITTLGTITATLSQ-----VINPAITVSAAGTTLTAAGGLTPTTMOV----- 954
QY 427 DHICKPVIGTPVQIKL-AQPGPVLSQP 452
Db 955 -----SOPTQVTLITAPSGVEAQP 973
RESULT 7
US-08-393-703-5
; Sequence 5, Application US/08393703
; Patent No. 5585239
; GENERAL INFORMATION:
; APPLICANT: Lamarco, Kelly
; APPLICANT: Wilson, Angus
; APPLICANT: Heit, Winship
; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
; TITLE OF INVENTION: HOST CELL FACTOR
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/393,703
; FILING DATE: 24-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57503-2/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-393-703-5
Query Match 9.1%; Score 204; DB 1; Length 2035;
Best Local Similarity 23.5%; Pred. No. 1.7e-08;
Matches 119; Conservative 66; Mismatches 188; Indels 134; Gaps 24;
QY 10 VSAPPKVSSGPRLPAPDIYAVKAPNTTITQPPANLQLPPTGVILIKSNGSLMLVSPQGV 69
Db 537 IGSSPQMSGAALAAAAAATOKIPSSA---PVLVSPAGTTIVKT-----MAVTPGTTT 588
QY 70 TRAETTSNISRPAVPANPQVTKICTVFNSSQLIKKVAIVPVKKLMDIGTTV----- 122

Db 569 LPA--TVAVASSPVAVNSPAT-----RMILKTA-----AAGTSVSSANTTS 628
QY 123 ---VTVPKPSSVOSVAAVTSVTVTPGKPLNTVTLK--PSSL--GASSTPSEPNL----- 173
Db 629 TRPITTVKSGTV--TVAQAQVTVTVVGVGVTKITILVKSPLSPVSGSALISNLKVMNV 687
QY 174 --KAENSAVQINLSPTMLENVKCKKNFL--AMLIKLACSSQSP-----EKQNVK 221
Db 688 QTKPVQTSVAVTGOASTGVTGTOIIOIKGPLPAGTILKLVTSADGKPTTITITVQASGACTG 747
QY 222 KLVEQLDAKLEAEFFTKLVELKSSPQPHVPLPKSVVALRQLNLSQSFLOQCQVQ 281
Db 748 PTLIGT-----SSVSPST---KPGTTITIKTIMSAITIQAGATG 785
QY 282 TSSDMVATCTTIVTT-----SPVVT-----TTVSSSQSEKSIIVSGA-----TA 321
Db 786 VTSSPGIKSPITITITTKMTSGTGAPAKIITAVPKIATIGHQOGVTVYVLGABQGPBTI 845
QY 322 PRT-----VSQTLNPLAGPVGAK--AGVTLHSVGPPTAAGTTAGTGLQTS 368
Db 846 LRTVPMGCVRLVPTVSAVAPVATTVLVKGTGTGTVTGTVST--SLAGAGHSTS 903
QY 369 KPIVTSVA--NTVTVTSIQPEKPVVSGTAVTSLPAVTFGETSGAALCLPSVKPVVSRGW 426
Db 904 ASLATPITTLGTITATLSQ-----VINPAITVSAAGTTLTAAGGLTPTTMOV----- 954
QY 427 DHICKPVIGTPVQIKL-AQPGPVLSQP 452
Db 955 -----SOPTQVTLITAPSGVEAQP 973
RESULT 8
PCT-US93-11721-5
; Sequence 5, Application PC/TUS9311721
; GENERAL INFORMATION:
; APPLICANT: Lamarco, Kelly
; APPLICANT: Wilson, Angus
; APPLICANT: Heit, Winship
; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
; TITLE OF INVENTION: HOST CELL FACTOR
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11721
; FILING DATE: 03-DEC-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: FP-57503-1/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide


```

APPLICANT: Dynlact, Brian D.
APPLICANT: Hoeey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,715
FILING DATE: 09-MAY-1996
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/188,582
FILING DATE: 28-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299-
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-646-715-2

Query Match 10.7%; Score 241.5; DB 1; Length 921;
Best Local Similarity 24.8%; Pred. No. 3.2e-12;
Matches 112; Conservative 65; Mismatches 163; Indels 111; Gaps 17;

QY 25 PQIVANKAINTTIOIPANLQIPGTVILKSNGLML-----VSFOQVYTRAETTS 76
DB 133 POSPTITLSTLNGOTPA-----LVKTDNGFQLRVGTTGPTVYOTITNTS 184
QY 77 NITSRAVPANPQTVKICTVPSNSSQ-----LIRKVAVPYKLAQIGTVVTV 127
DB 185 NITSTTNHFTTQ-IRLOVPPAAASMTNTATSNITVNSVASSGYANSQPPHILQ 243
QY 128 KSSVQSAVPTSVTVTGKPLNTVTLKPSLSGASSTPSENPILKAENSAVAQINLSP 187
DB 244 APQLPQITQITIPAOSSQOQVNNVSSAGATATAVSSFTA-----ATT 287
QY 188 TULEANK-KCKNPLMLILACSGSPPMGQNVKLVQQLDAKLEAEFTKRLVEIK 246
DB 288 TOGNTKKECKRFLANLIL-STRPKPEKRVRLTLOELVNAVPEPEFCDRLERLIN 345
QY 247 SSPPHLVEFLKRSVALROL-----LPNSQSFIO--- 276
DB 346 ASBPOLIGFLKSLPLRLQALYKELVLEGKPPROHVLGLAGLSQOLPKIOAQIRPIG 405
QY 277 ---QCVOQTSSDMVIACTTTVTTSPTVTVSSQSEKSIIVSGATAPRTVS---VOT 329
DB 406 PSQTTIGQTVRMI--TFNALGTPTPTIGHTTISKOPN---IRLPTAPRLVNTGIR 460
QY 330 LNPAGPVAKAGVYTLHSGVTPAATGTTAGTGLQTSKPLVTSVANTVTYVSIQPEK 389

```

```

DB 461 QIP-SLOVPGQANIVQIR--GPOHAQLORTGVSQVIRATTRP-----PNSVPTAN----- 506
QY 390 VVSGTAVTLSPRVTFGTSGAALICPSVKP 420
DB 507 -----KLTAVKVQGTQIKAI-TPSLHP 527

RESULT 6
US-08-046-585-5
Sequence 5, Application US/08046585
Patent No. 5453362
GENERAL INFORMATION:
APPLICANT: Lamarco, Kelly
APPLICANT: Wilson, Angus
APPLICANT: Herr, Winship
TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
TITLE OF INVENTION: HOST CELL FACTOR
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/046,585
FILING DATE: 12-APR-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57503-1/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299-
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-046-585-5

Query Match 9.1%; Score 204; DB 1; Length 2035;
Best Local Similarity 23.5%; Pred. No. 1.7e-08;
Matches 119; Conservative 66; Mismatches 188; Indels 134; Gaps 24;

QY 10 VSAIPVSSGRLPAPQIVAVKAPNTTIOIPANLQIPGTVILKSNGLMLVSPQTV 69
DB 537 IGSSPPMSGMALAAALAAATOKRIPSSA---PVLVSVPAGTTIVK-----NAVITGITT 588
QY 70 TRAEITSNITSRAVPANPQTVKICTVPSNSSQLIRKVAVTPYKLAQIGTV----- 122
DB 589 LPA--TVKVAASSPVAVNSPAT-----RLKTA-----AQVGTSVSATNTS 628
QY 123 ---VTTPKPSVQSAVPTSVTVTGKPLNTVTLK-PSST-GASSTPSEPNL----- 173
DB 629 TRPITVHRSQV-IVAQAQAVTVVGGVTRTITLVKSPISVPGSALISNIGKVMSTV 687
QY 174 --KAENSAVAQINLSEPTMLENKKCNFL--AMLKILACSGSQSP-----EMQONVK 221
DB 688 QIKPVQTSANVTOASGTPYTOIQRKPLPACTIILKLVTSADAGKPTTITTTTQASAGAK 747
QY 222 KLVEQLDAKLEAEFTRLRYVELKSSQPHLVPLKRSVALROLPLNSQSFIOQVQO 281

```


Db 61 MLVSPQIVTTRAEITNSITSRPAVPANPQTVKICIVPNSSSOLIKKVAATPVKKLAQIGT 120
QY 121 TVTTPKPSVQSVAVPTSVVTPGKPLNTVTTLKPSLSGASTPSNEPNKAENSA 180
Db 121 TVTTPKPSVQSVAVPTSVVTPGKPLNTVTTLKPSLSGASTPSNEPNKAENSA 180
QY 181 VOINISPTMLENVKCKNFKNLAMLIKACSGSOSPENGONVKIYVOLLDAKIEAEFTTR 240
Db 181 VOINISPTMLENVKCKNFKNLAMLIKACSGSOSPENGONVKIYVOLLDAKIEAEFTTR 240
QY 241 LVEYLKSSPOPHLVPFLKSVVALROLPLNSOSFIOOCVOQTSDDVIACTTTVTTSVP 300
Db 241 LVEYLKSSPOPHLVPFLKSVVALROLPLNSOSFIOOCVOQTSDDVIACTTTVTTSVP 300
QY 301 VTTTSSSSSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVYTLHSVGTATGTTA 360
Db 301 VTTTSSSSSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVYTLHSVGTATGTTA 360
QY 361 GTGLQTSKPLVTSVANVTTVSVLSQPEKPVSGTAVTSLPAVTEGTSGLAICLPSVR 420
Db 361 GTGLQTSKPLVTSVANVTTVSVLSQPEKPVSGTAVTSLPAVTEGTSGLAICLPSVR 420
QY 421 VVSECDHICKPIGTPVOIKLAQPGPVLSQP 452
Db 421 VVSECDHICKPIGTPVOIKLAQPGPVLSQP 452

RESULT 2

US-08-188-582-16
Sequence 16, Application US/08188582

GENERAL INFORMATION:
APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dynlacht, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS.
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAPS AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESSES:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,582
FILING DATE: 28-JAN-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 737 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-188-582-16

Query Match 23.2%; Score 521; DB 1; Length 737;
Best Local Similarity 34.0%; Pred. No. 2, 2e-36;
Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

QY 13 PPKVSSG-----PRLPAQIVAAKAPNTTTIQEPANILQPLPGTVLKSNGLM 61
Db 68 PRATTSIGRATLPTVIALPRLPQ-----PQNPINIQ---NFQLPGVNLVSENGOLL 118
QY 62 LVSPOQTIVR-----AETTSNITSRPAVPANPQTVKICIVPNSSSOLIKKVAATPVKKL 115
Db 119 MI-PQALAQMOQAHAQOQTMAAPPAAPTSAPVQISTVQAAGPPIAR-QVTP----- 172
QY 116 AQIGTVTTVPKPSVQSVAVPTSVVTPGKPLNTVT--TLKPSLSGAST-----TPS 168
Db 173 ----TTIIKQV---SQAQTTVPASATLQSPGVQPOLVLGGAQVTSLSGTATVQTGTQ 225
QY 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFKNLAMLIKACSGSOSPENGONVKIYVOLL 227
Db 226 RVPGATTTSSATE-----TWENYKCKNFKNLSTLIKLASGKSTETAAVYKELVQML 279
QY 228 LDKAIEAEFTKRLVYELKSSPOPHLVPFLKSVVALROLPLNSOSFIOOCVOQ----TS 283
Db 280 LDKAIEAEFTKRLVYELKSSPOPHLVPFLKSVVALROLPLNSOSFIOOCVOQPPPTS 339
QY 284 SDMVIACTTTVTSPVTTTSSSSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGV 343
Db 340 Q-----ATTALTAVYLVSSVQRTGKTAAYTSLQPPVLSL----- 376
QY 344 VTLHSVPTAAGTGTAGTGLQTSKPLVTSVANVTTVSVLSQPEKPVSGTAVTSLPAV 403
Db 377 -----TQPTQVGVKGQGPPLVLIQ-----QPPKP-----GALIRPQV 410
QY 404 TREGTSGLAICLPSVAPVSVFCDHICKPIGTPVOIKL--AQPGVYL 449
Db 411 TLITQT-----PMVALRQPH-NRIMLTTPQIDQLNPLOPVV 446

RESULT 3

US-08-646-715-16
Sequence 16, Application US/08646715

GENERAL INFORMATION:
APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dynlacht, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS.
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAPS AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESSES:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,715
FILING DATE: 09-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/188,582

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:59:29 ; Search time 11.6146 seconds
(Without alignments)
1145.042 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_452
Perfect score: 2249
Sequence: 1 GTLVTKVAPVAPVSSGP.....VIGTPVQIKLAGPVLSP 452

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCVUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2249	100.0	801	1	US-08-725-012-2
2	521	23.2	737	1	US-08-188-582-16
3	521	23.2	737	1	US-08-646-715-16
4	241.5	10.7	921	1	US-08-188-582-2
5	241.5	10.7	921	1	US-08-646-715-2
6	204	9.1	2035	1	US-08-046-585-5
7	204	9.1	2035	1	US-08-393-703-5
8	204	9.1	2035	5	PCT-US93-11721-5
9	176.5	7.8	2972	4	US-09-579-181-2
10	176.5	7.8	3118	4	US-09-579-181-1
11	172	7.6	750	4	US-09-165-239A-4
12	171.5	7.6	1721	3	US-08-700-651-5
13	171.5	7.6	1721	3	US-08-928-361B-6
14	164.5	7.3	752	1	US-08-244-189-2
15	163.5	7.3	862	1	US-08-325-267A-4
16	161.5	7.2	1837	3	US-08-928-361B-5
17	159.5	7.1	894	3	US-08-362-525-22
18	159.5	7.1	894	3	US-08-971-692-15
19	156	6.9	941	4	US-07-757-022B-14
20	156	6.9	1022	4	US-07-757-022B-84
21	156	6.9	1038	4	US-07-757-022B-74
22	156	6.9	1049	4	US-07-757-022B-58
23	156	6.9	1140	4	US-07-757-022B-104
24	156	6.9	1270	4	US-07-757-022B-44
25	156	6.9	1311	4	US-07-757-022B-42
26	156	6.9	1313	4	US-07-757-022B-142
27	156	6.9	1314	4	US-07-757-022B-50

28	156	6.9	1320	4	US-07-757-022B-46	Sequence 46, Appl
29	156	6.9	1320	4	US-07-757-022B-60	Sequence 60, Appl
30	156	6.9	1354	4	US-07-757-022B-48	Sequence 48, Appl
31	156	6.9	1361	4	US-07-757-022B-40	Sequence 40, Appl
32	156	6.9	1363	4	US-07-757-022B-52	Sequence 52, Appl
33	156	6.9	1404	4	US-07-757-022B-2	Sequence 2, Appl
34	156	6.9	1404	4	US-07-757-022B-62	Sequence 62, Appl
35	155	6.9	907	3	US-08-783-774-2	Sequence 2, Appl
36	155	6.9	907	5	US-09-328-559A-1	Sequence 1, Appl
37	155	6.9	907	5	PCT-US95-04611A-19	Sequence 19, Appl
38	153	6.8	878	4	US-09-556-706B-2	Sequence 2, Appl
39	150	6.7	1537	1	US-08-325-267A-2	Sequence 2, Appl
40	148.5	6.6	805	4	US-09-103-429A-4	Sequence 4, Appl
41	148	6.6	786	4	US-09-103-429A-3	Sequence 3, Appl
42	146.5	6.5	3969	4	US-08-061-376-5	Sequence 5, Appl
43	142.5	6.3	2476	2	US-08-276-967-2	Sequence 2, Appl
44	140	6.2	1147	1	US-08-131-365B-38	Sequence 38, Appl
45	140	6.2	1147	2	US-08-668-123-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1
US-08-725-012-2
Sequence 2, Application US/08725012
Patent No. 5710025
GENERAL INFORMATION:
APPLICANT: Dikstein, Rivka
APPLICANT: Tjian, Robert
TITLE OF INVENTION: B-cell Specific Transcription Factor
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/725,012
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: B97-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 801 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-725-012-2
Query Match 100.0%; Score 2249; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 7.9e-185;
Matches 452; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GTLVTKVAPVAPVSSGPRLPAPQIVAKAPMTTITQFPANTQLPPTGLVLIKNSGFL 60
DB 1 GTLVTKVAPVAPVSSGPRLPAPQIVAKAPMTTITQFPANTQLPPTGLVLIKNSGFL 60
OY 61 MLVSPQGVTAETSNITSPAPVAPNQTWKICTVPRSSQLKKAVTVYKKLAQIGT 120

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:31:01 ; Search time 31.7465 Seconds
(without alignments)
2933.659 Million cell updates/sec

Title:	US-09-763-909-2_COPY_1_452
Perfect score:	3249

Sequence: 1 GTLVTKVAPVSAPPKVS...VIGTPVQIKLAQPGPVLSP 452

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database : SPTREMBL_21:*

```

1: 1: sp_archaea:*
2: 2: sp_bacteria:*
3: 3: sp_fungi:*
4: 4: sp_human:*
5: 5: sp_invertebrate:*
6: 6: sp_mammal:*
7: 7: sp_mhc:*
8: 8: sp_organelle:*
9: 9: sp_phase:*
10: 10: sp_plant:*
11: 11: sp_podent:*
12: 12: sp_virus:*
13: 13: sp_vertebrate:*
14: 14: sp_unclassified:*
15: 15: sp_virus:*
16: 16: sp_bacterioph:*
17: 17: sp_archaeoph:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	526	23.4	662	11	Q91W6	Q91W6 mus musculus
2	220	9.8	851	5	Q8T9E0	Q8T9E0 drosophila
3	215	9.6	1275	5	Q766D0	Q766D0 caenorhabdi
4	215	9.6	1349	4	Q8WMO4	Q8WMO4 homo sapien
5	209.5	9.3	528	6	Q29071	Q29071 sus scrofa
6	205	9.1	1979	11	Q9OY35	Q9OY35 mus musculus
7	202	9.0	1795	5	Q76894	Q76894 drosophila
8	200	8.9	2445	11	Q61191	Q61191 mus musculus
9	198.5	8.9	842	5	Q9VGC9	Q9VGC9 drosophila
10	199.5	8.9	864	5	Q95U45	Q95U45 drosophila
11	199.5	8.9	864	5	Q9VGC8	Q9VGC8 drosophila
12	198	8.8	513	4	Q43418	Q43418 homo sapien
13	198	8.8	2045	11	Q9QWH2	Q9QWH2 mus musculus
14	197.5	8.8	1322	4	Q8T9E0	Q8T9E0 homo sapien
15	195	8.7	1079	5	Q9N4S7	Q9N4S7 caenorhabdi
16	192	8.5	1844	5	Q22579	Q22579 caenorhabdi

17	191.5	8.5	1893	5	Q9NKC9	O9pkc9 drosophila
18	191	8.5	1246	4	O15052	O15052 homo sapien
19	191	8.5	2382	4	O9H4A3	O9H4A3 homo sapien
20	187	8.3	1029	4	O8TDH7	O8TDH7 homo sapien
21	186.5	8.3	873	10	O9C548	O9C548 arabidopsis1
22	186.5	8.3	5374	11	O99ND0	O99ND0 mus musculi
23	186	8.3	389	12	O8V0M0	O8V0M0 equine herra
24	185.5	8.2	3570	4	O99552	O99552 homo sapien
25	185.5	8.2	3971	3	O96W66	O96W66 schizosacch
26	185	8.2	648	4	O14760	O14760 homo sapien
27	183.5	8.2	2232	5	P91365	P91365 caenorhabdi
28	183	8.1	800	3	O8TF64	O8TF64 schizosacch
29	182.5	8.1	374	12	O8V0L6	O8V0L6 equine herra
30	182.5	8.1	2187	11	P70670	P70670 mus musculi
31	181.5	8.1	356	12	O8V0L7	O8V0L7 equine herra
32	181.5	8.1	709	4	O9NVJ9	O9NVJ9 homo sapien
33	181.5	8.1	961	3	O92223	O92223 emericella
34	179.5	8.0	645	4	O9H048	O9H048 homo sapien
35	179.5	8.0	867	12	O39782	O39782 equine herra
36	179.5	8.0	1324	4	O15312	O15312 homo sapien
37	178.5	7.9	622	4	O14881	O14881 homo sapien
38	178.5	7.9	1716	11	O9JK31	O9JK31 mus musculi
39	178.5	7.9	1301	5	O8TK34	O8TK34 drosophila
40	178	7.9	886	12	O8V0L5	O8V0L5 equine herra
41	178	7.9	886	12	O9Q087	O9Q087 human herpe
42	177.5	7.9	662	5	O9VG00	O9VG00 drosophila
43	177.5	7.9	3507	5	O23587	O23587 caenorhabdi
44	177	7.9	1236	3	O9C105	O9C105 schizosacch
45	176.5	7.8	1203	5	O9NSK0	O9NSK0 caenorhabdi

01	091MW6	PRELIMINARY;	PRF:	662 AA.
02	091MW6			
03	01-DEC-2001	(TREMBLrel. 19, Created)		
04	01-DEC-2001	(TREMBLrel. 19, Last sequence update)		
05	01-JUN-2002	(TREMBLrel. 21, Last annotation update)		
06	TARF-binding protein associated factor TAFII135 (fragment).			
07	TAF4A.			
08	Mus musculus (Mouse).			
09	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
10	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
11	NCBI_TaxID=10090;			
12	[1]			
13	SEQUENCE FROM N.A.			
14	STRAIN-BALB/C; TISSUE-PO BRAIN;			
15	Metsis M., Brunkhorst A., Neuman T.;			
16	"Cell Type Specific Expression of the TrlD Component TAFII135 in the			
17	Nervous System.";			
18	Exp. Cell Res. 0:0-0(2001).			
19	EMBL; AY038601; AAK94779.1; -.			
20	MGD; MGI:2152346; Tafa4a.			
21	NON-TER	1		
22	SEQUENCE	662 AA;	71398 MW;	91A75F38CB0D0DA4 CXC64;
23	Query Match	23.4%;	Score 526;	DB 11;
24	Best Local Similarity	35.0%;	Pred. No. 3.5e-25;	
25	Matches 155;	Conservative 60;	Mismatches 128;	Indels 100;
26				Gaps 18;
27	20	PLPLPQIVAVVAPNTTITPPANLQLPFGVILIKNSGSLMLVSPOQTV-----TRA	73	
28	10	PLPLPQ-----QKPL-NIQ--NFQPLGMYLHSENGQLMTI-PQQLAQMOMHAQMQ	59	
29	74	TTSNTSSRPANPQTVKICIVPPSSSSOLIKKIVAVTEPVKKLQIGTIVTTVPKPSVQ	133	
30	Db	60	PGSTMAPRPAHPITGAPVQISTVQAPGPIIAR-QVTP-----TTIKVY--SQAQ	107
31	09	134	SVAVETSVVYVTPGKPLMTVT--TIKPSLSGAS-----TPSNE-PNKAENSAAVQINTL	185

Db 108 TVVOPPTTLQSRSPGVQPOLVIGSQAQASIGATATAVGTGTPORTVPGASTTSTATE---- 164
 Qy 166 SPTMLENKKCNFLAMLKILACSGSQSPKONKVLLEQLLDKATEEETRLKLYEL 245
 Db 165 ---TBNENKKSFLSTLLKSSGQSTETANVDLQNLIDGTEAEDFTSRLRYEL 221
 Qy 246 KSSPQHLVFPFKSSVALROLPLNPSQSFIOOCVOOTSQSDMTATCTTPTVTSPTVTT 305
 Db 222 NSSPQPLVFLFKRSIPALROLPLPSAATIOSQOQPP-----ASQATTLALAVLSSV 277
 Qy 306 SSSQSEKSIISGATAPRTVSQTLNPLAGPVAKAGVTLHVSFPPTAGTTACTGL 365
 Db 278 QRTAGKTASVTSALQPPVISTL-----TQPTQVGVGKQ 310
 Qy 366 QTSKPLVTSVANTVTTSLSIQPEKPVVSGTAVTLSPATVTFGEGSCAICLPKPVVSPFC 425
 Db 311 APPTPLVIG-----QPPK-----GALIRPQVTLTQT-----PVALR 344
 Qy 426 WDICKPVIPTVQIKLAPGPV 448
 Db 345 QPH-NRIMLTTPQIQLOLQOPV 366

RESULT 2

Q8T9EO PRELIMINARY: PRT: 851 AA.
 AC Q8T9EO:
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE SD04735P.
 GN TAF10.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 OX [1]
 RN RP
 RA SEQUENCE FROM N.A.
 RA Streptococcus M., Brockstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farhan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Lao G., Miranda A., Mungall C.J.,
 RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
 RA Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AY069807; AAL39952.1;
 SO SEQUENCE 851 AA; 92093 MW; 665B28B9586C984C CRC64;

Query Match 9.8%; Score 220; DB 5; Length 851;
 Best Local Similarity 26.5%; Pred. No. 7.7e-06;
 Matches 117; Conservative 59; Mismatches 146; Indels 120; Gaps 22;

Qy 22 LPAPQIVAVK--APNTTITQ--FPANLQLPPTGVILKSNGLMLVSPQOTVRAETTSNI 78
 Db 93 LPAGVAVGKROQAPSOQOKNMPFN---PLSRVIVINSHAGVROPSP--SKMTNTTATSI 148
 Qy 79 TSRAVAVANPOTVXICTVPSNSS--QLIKKVAVTVPYKKAQIGTVVTVTPKPSVQSA 136
 Db 149 I---VNSVASSGTANSSQPHLQOLNAOP--QLPQI--TQIQITPAAQSOQ--- 193
 Qy 137 VPTSVVTVTPGKPLNTVTLTKPSSLGASSTPSNEPLKAENSAVAOINLPTMLENVK-K 195
 Db 194 ---QOVNNSVAGSTATAVASTTA-----ATTTOQGQTKRK 226
 Qy 196 CKNFLAMLIKILACSGSQSPKONKVLLEQLLDKATEEETRLKLYELKSSPQHLVP 255
 Db 227 CRKFLANIEL--STRKRPVKKNVRTILOELVNNANVEEFCDELRRLNLSPOCLIG 284
 Qy 256 FLKKSVALROL-----LPSQSFIO-----QCVOQ 281
 Db 285 FLKKSVALROL-----LPSQSFIO-----QCVOQ 281
 Qy 282 TSSDMVATCTTPTVTSPTVTSVSSQSEKSIISGATAPRTVS---VQTLNPLAGPVG 338

Db 345 TOYRMI--TPNALGTRPPIIGHTTISKOPPN---IKLPAPRLVNTGCTQIIP-SLOVP 398
 Qy 339 AKAGVTLHVSFGPPTAGTTACTGLQTSKPLVTSVANTVTTSLSIQPEKPVVSGTAVTL 398
 Db 399 GQANIYQIR--GPQHQLQRTGTSVQIRATTRP-----PNSVPIAN----- 436
 Qy 399 SLPAVTFEGESGAICLPKVP 420
 Db 437 KILAVKVGQTLKAI--TPSLHP 457

RESULT 3

OT6602 PRELIMINARY: PRT: 1275 AA.
 AC OT6602:
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Hypothetical 122.9 kDa protein.
 GN H02F09.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabditoidea;
 OC Rhabdilitae; Peloderinae; Caenorhabdilita.
 OX NCBI_TaxID=6239;
 [1]
 RN RP
 RA SEQUENCE FROM N.A.
 RA STRAIN-BRISTOL N2;
 RX MEDLINE-94150718; PubMed-7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berke M.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favejlo A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,
 RA Smaildon N., Smith A., Sonnenhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Spoat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RN Nature 368:32-38(1994).
 [2]
 RN RP
 RA SEQUENCE FROM N.A.
 RA STRAIN-BRISTOL N2;
 RA Geisel C., Harmon G.;
 RT "The sequence of C. elegans cosmid H02F09.";
 RT Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
 [3]
 RN RP
 RA SEQUENCE FROM N.A.
 RA STRAIN-BRISTOL N2;
 RC Waterston R.;
 RL Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF07538; AAC64622.1;
 KW Hypothetical protein.
 SO SEQUENCE 1275 AA; 122924 MW; CB59B70C05959E25 CRC64;

Query Match 9.6%; Score 215; DB 5; Length 1275;
 Best Local Similarity 23.6%; Pred. No. 2.6e-05;
 Matches 110; Conservative 71; Mismatches 182; Indels 104; Gaps 18;

Qy 2 TLVTKAPSPAPKVSSEF---RLPAPQIVAVKAPNTTITQFPANLQLPPTGVILKSN- 56
 Db 275 TVVPTVTVTPGPTVTVVTPVTVVTPVTSPTVTSPTVTSPTVTVVTPVTVVTPVTVVTPVTSPT 334
 Qy 57 ---SGPLMVS--POQVTVRAETTSNITSRAVAVANPOTVXICTVPSNSSQLIKKVAANTPVK 113
 Db 335 VVTAPSTVTVSTVTVTKNTV--VTSPTVATPTTV--VTPPT-----VTVPS 382
 Qy 114 KLAQIGTIVV---TVVPPSSVQSAVAPTSVT-----VTPGKPLNTV 153
 Db 383 TVVTVVTVTVVTPVTVVTPVTVVTPVTVVTPVTVVTPVTVVTPVTVVTPVTVVTPVTVVTPVTVVTP 440


```

OY 271 SOS-----FIQ-----QCYOQSSDMVIANCTTVMTS-----PVYTTTV 305
DB 360 SSSSSSTTTSTTSVQPSSSSGAPTTSATSVQPSSSSSVPTTTSATSVSSSSSPPTPTTTS 419
OY 306 SSSOSEKSIIVSGATAPRTVSVQTLNPLAGPVAKAGVTLHSGVTAATG--TIAAGTGL 364
DB 420 VQPSSSSSSVPTTTSAT-----SVQTSSSSSSPRTSTT-----SVQPSSSSSATTTATSV 468
OY 365 LOTS-----KPLVTSVANTVTTVSLOPEKPVVSGTAVTSLPAVTFEGTSGAA 412
DB 469 QPSSSSSPPTSTTSVQPSSSSSPTTSTTSVQPS-----SGSAPTTSATSVQPSSSSP 525
OY 413 I 413
DB 526 I 526

RESULT 6
OYQY35 ID 090Y35 PRELIMINARY: PRT: 1979 AA.
AC 090Y35: 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Host cell factor C1 (Fragment).
GN ABCD1 OR HCFC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Crustacea; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Platzner M., Brenner V., Reichwald K., Wiehe T., Oksche A.,
RA Rosenbluth A.;
RT "Comparative sequence analysis of the mouse Licam locus and the
RT corresponding region of human Xq28."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF133093; AAF2156.1;
DR MGD: MGI:105942; Hcfcl.
DR MGI:1349215; Abcdl.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF01344; kelch; 4.
DR SMART: SM00060; FN3; 1.
FT NON-TER 1
SQ SEQUENCE 1979 AA; 203436 MW; ABE998C3DDE688A9 CRC64;

Query Match 9.1%; Score 205; DB 11; Length 1979;
Best Local Similarity 23.7%; Pred. No. 0.00019;
Matches 128; Conservative 63; Mismatches 209; Indels 140; Gaps 26;

OY 4 VTKVAVSAPKPVSGPRL--PAPQIVAKAPNTTIOFPANL-----QLPPTGTVL-- 52
DB 417 VPAVLKVTGPQATGTPPLVTPMPASQ--AGKAPVTV--SLPASVAVVPLQSGAQVIGS 473
OY 53 -----IKNSGPLMVSPQGYTRAEFTTSNITSRPAVPANPQTVKI 93
DB 474 NPQMSGMAALAAAAATQKIPSSAPTVLSVPAGIT-----IVKTVAVTGGTTLLPATVVK 529
OY 94 CTVP-----NSSQLIKKVVV--TPVKLAQIGTVVTPKPSQSVAVPVSQVAVP 146
DB 530 ASSPVAWSAPATRMKLTAAAOQVTSVSSAANSTRPITVHKSQV--TVAAQAOVTVTV 568
OY 147 GRPLNTVTTLK--PSSL-GASSPSPNEPNL-----KAENSAVQVIMLSPMLENKKCN 198
DB 589 GGVTKITLVKSPISVPGSALISNIGKVMVSVQKPVQTSVATGAGSGVPIVQIQTG 648
OY 199 FL--ZMLITLACSGSQSP-----EMQONKYLVEQLIDAKIEEFTTKLYVELKSS 248
DB 649 PLPAGTITLKVTSADKPTTITTTTQASGAGTKPTILGI-----SS 689
OY 249 PDPHLVPLFKKSVVALRLQLPNSQSFIOQVOQTSSDMVIANCTTTVT-----S 298

```

```

DB 690 VSPSTT---KPGTTTTIKTIPMSAIIITQACAGTGVSSPGIKSPITTTTVMVMTSGTAPA 746
OY 299 PVVT-----TVSSOSEKSIIVSGA-----TPRT-----VSQVTLNPLAGPV 337
DB 747 KITVAVKIKTIGGQGVNOVLKAGAPGCRPTILRTVPMGVALYMPVTVSAKPAVATL 806
OY 338 GAK--AGVVTLHSGVPLAATGCTTACTGLQTSKPLVTSVA--NVVTVSLOPEKPVSG 393
DB 807 VVKGTGTGVTIGTGTGVST--SLAGAGHSTSAASLATPTTIGTITATLSSQ-----VINP 860
OY 394 TAVTSLPAVTFEGTSGAATCLPSVKPVVSEFCDHICKPVGIVQOKL--AQGGPVLSQP 432
DB 861 TAITVSAQTTTLTAAGGLTPTTTPMPV-----SQPTQVTLTAAPSVEAQP 907

RESULT 7
OYQY35 ID 076894 PRELIMINARY: PRT: 1795 AA.
AC 076894: 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE EG:56G7.1 protein.
DE EG:56G7.1 OR CG14796.
GN EG:56G7.1 OR CG14796.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-BERKELEY;
RC MEDLINE-20196006; PubMed-10731132;
RX ADAMS M.D., Cealiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amandlides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Fandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
RA Adair J.F., Aspayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Broxstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davoport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Fostler C., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,
RA Jalali M., Kalush F., Kapran G.H., Ke Z., Kenison J.A., Kethum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Put V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spletas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasthman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]

```



```

RP SEQUENCE FROM N.A.
RA Cadien E., Dreano S., Lelaure V., Motlier S., Gallibert F.;
RT "Sequencing the distal x chromosome of Drosophila melanogaster.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Benos P.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AEO03421; AAF4564.1; -
DR EMBL: AL031028; CA19845.2; -
DR FlyBase: FBgn0025390; EG:5667.1.
DR InterPro: IPR002557; Chitin_bind_Pera.
DR InterPro: IPR002965; P-rich_extensn.
DR Pfam: PF01607; CBM_14; 2.
DR PRINTS: PR01217; PRICHEXTENSN.
DR SMART: SM00494; ChIBD2; 2.
SO SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;

Query Match 9.08; Score 202; DB 5; Length 1795;
Best Local Similarity 21.88; Pred. No. 0.00026;
Matches 95; Conservative 58; Mismatches 196; Indels 86; Gaps 13;

QY 9 PVSAAPKVSQGR.LPAPQIVAVKAPNTTIOFANLQI-----PGTVLTKNSG 58
DB 700 PPSSTGPTTPK-PSRTPTPTTKVITITQITTPPLRSTETTSQPPPTTPPQPTT 758
QY 59 PLMLVSPQOQVTRAEETSNITSRPVAPNPQVTKICVNPSSQILKKVAVTPVKLAQI 118
DB 759 TTLVTFKTSVTYTTTEKPIITSPK-PYTTQKTSTAPNNT-----KVATITQKPTPT 812
QY 119 GTT-----VTVTPKPS-----VQSAVPPSVTVTPGKPLNT 152
DB 813 QSTSTFTTTRKTTNNPEPTSTEKPIITSTPKPSTTPKSTVASTEKTISSPKPTTE 872
QY 153 VTTLPSSLSASSPSPNEPNLKAENSAVQINISPTMLENKKKKFLAMIKILACSGSQ 212
DB 873 KSTNPPTNSYKISALSSSTORA-----TSTSEPTKTQNTTTTPKPTTKTS 922
QY 213 SPENGQVKKLVLEOLDLAK--IEAEFTFKLVLELKSSPOHLPFLKKSVAALROLIPN 270
DB 923 TQEAFTTSQKSVYITTKKATESPLTTLSTEBNTPPK-----LRTITPT 970
QY 271 SOSTQOCVOQTSNDVIATCTTVTTSPPVYTTVSSSQSEKSIIVSGA--TAPR--TVS 326
DB 971 TTS-----VYATTRITTTTISSESTETSTQKPSSTPSTPTKPTKVTYVI 1017
QY 327 VQTLNPLAPGKAGAVVTLHSVGPFAATGTFAGTGLQTSKPLVSVANVTTVSLQP 386
DB 1018 VSTQNPPT--TTSKTSVTITTPNPSPSTORPTTTTQPTSTASTSIGTRIPPTTN 1075
QY 387 EKPVSCTAVTSLP 401
DB 1076 QNSTSTDLTPTTRP 1090

RESULT 8
Q61191 PRELIMINARY; PRT; 2045 AA.
AC 061191;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Transcription factor C1 (HCF).
GN HCF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TTSUE-LIVER;
RA Kissile T.M., Dasher R.;
RT "CDNs encoding the mouse homolog of the human transcription factor C1

```

```

RT (HCF).";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U53925; AAB01163.1; -
DR MGD: MGI:105942; Hcf1.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR001796; Kelch.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF01344; Kelch; 5.
DR SMART: SM00060; FN3; 1.
SO SEQUENCE 2045 AA; 210537 MW; B89CC2FDA35B969F CRC64;

Query Match 8.98; Score 200; DB 11; Length 2045;
Best Local Similarity 24.18; Pred. No. 0.0004;
Matches 129; Conservative 68; Mismatches 201; Indels 138; Gaps 28;

QY 7 VAPVAPKVSQGR.LPAPQIVAVKAPNTTIOFANLQI.LPBGVULKNSGPLMLV 63
DB 486 VLKVTGPQATGTGPLVTRPQSO--AGKAPTVV-SLPASVPM--VPTQSAQGTIVGS 539
QY 64 SPQQT-----VTRAEETSNI--TSRP--AVPAN-----PQVTKICTVP 97
DB 540 NPQMSGMAALAAATAQKIPSSAPITAMSVPACTTIVKTVAVTPGTTLPATVAVASSP 599
QY 98 -----NSSQILKKVAV--TPVKLAQIGTTVTVTPKPSVQSAVAVPTSVTVTPGKPL 150
DB 600 VMVSNPATRMLKTAAGVGTSSVSSAANTSFRTITVHKSGTV--TVAAQAVTVTVGCVT 658
QY 151 NTVTTLK-PSL-GASSPSPNEPNL-----KAENSAVQINISPTMLENKKKKNFL-- 200
DB 659 KTTLVKSPISVPGSALISMGRKMSVQOTKPVQTSVAVTQASTGPTQITQKGPLA 718
QY 201 AMLIKLACSGSQSP-----EMQNVKKLVLEOLDLAKIEAEFTFKLVLELKSSPOH 252
DB 719 GTLIKIVTSADCKPTTITTTQASGAGTKPILGI-----SSVSPS 759
QY 253 LVPLKKSVALROLPLNSOSFIQOCVOQTSNDVIATCTTVT-----SPVY 302
DB 760 TT---KPGTTTIIKITPMSALITQAGATGVTSSPGKISPIITITTKVTSQTPAKIT 816
QY 303 -----TVSSSQSEKSIIVSGA--TAPR-----VQTLNPLAPGK- 340
DB 817 AVPKIATGGOQGVQVYVLAAGAPGPGTILKTVPMGVRLVTPVAVPAVTVLVK 876
QY 341 AGVTLHSVGPFAATGTFAGTGLQTSKPLVSVN--NTVTVSLQEPKPVSGTAVT 397
DB 877 TTGVTITIGTVIGVST--SLAGAHSTASLAPITTLGIALLSQ-----VINPAIT 930
QY 398 LSPAVTFGETSGAALCLPSVKPVVSCMDHICKPVIQTVQIKL-AQPGVLSQP 452
DB 931 VSAQOTLTLAAGLTPPTITWQPV-----SQPTQVTLTAPSGVEAQP 973

RESULT 9
Q9VGC9 PRELIMINARY; PRT; 842 AA.
AC 09VGC9;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE CPN protein.
GN CPN OR CG4795.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

```

RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Mikos G.L.G.,
 RA Abriil J.F., Agbayani A., An H.-J., Andrews-Plankoch C., Baldwin D.,
 RA Bailett R.M., Basu A., Bakendale J., Bayraktaroglu T., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,
 RA Jalali M., Kalush F., Karpen G.T., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshirei A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitzkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Massarman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RA Science 287:2185-2195(2000).
 RA EMBL: AE003694; AAF54755.1; -
 RA Flybase: FBgn0010218; Cpn.
 RA InterPro: IPRO02965; P-rich extensin.
 RA PRINTS: PR01217; PRICHEXTENSIN.
 RA SEQUENCE 842 AA; 82242 MW; D71E531327EF8501 CRC64;

Query Match 8.9%; Score 199.5; DB 5; Length 842;
 Best Local Similarity 22.5%; Pred. No. 0.00015;
 Matches 124; Conservative 70; Mismatches 203; Indels 153; Gaps 25;

QY 1 GIVTVTV-APVSA-----PRVSGPRLPAP---QIVAKAPNTTIOFPANIQ 45
 DB 4 GIIPEVSAAPVAAVTPPSAAVAPVQVSPAAVAPAPAAPIAVTPVAPPTLASVQATAT 63
 QY 46 LPPGTVLINSNGPLMLVSPQOTVTRAETTSNTTSRPAPVAPANQVYKICVPPNS-SSQLI 104
 DB 64 IPAPAPIAASVAPVAVSAP--PVVAAPTPP-----NASVSTPPVAVAOIPVAVASAPVA 116
 QY 105 KRYAVT-----PVKKLAQIGTT--VVTIVKPPSSVOSVANPTSVVTVTGKPLNTTTLK 157
 DB 117 PVAATPTTPVAPVAPVATPPVAAAPVAPVAVTPVAVPVPVSP--VIAITPPVANTTV--- 171
 QY 158 PSSLGASSTPSPNEPNKKAENSAVQINLSPTMLE-----NKKCKNFLAMLI 204
 DB 172 PVAAPVAAVAPVAVPVLAPVAPVAVVAVETPAPPVAEIPATIPREC---AAPLI 228
 QY 205 KLASGSGSPKMGONKVLVEQLDAKIEAEETTRKLYELKSSPPHVLVPLFKKSVAL 264
 DB 229 -----PEVSVATKPLAAAEPPVAVAPPATETPVVAPAAASPVAVAPVAVETAIVA- 278
 QY 265 ROLPNSOSFIOQCVOOTSSDMVIATCTTWTTSPPVTTTSSSOSSEKSIIV-----SGAT 320
 DB 279 -----PVASAS-----TTPVAAATLTITAPET--PALAPVAVESQVAAANTVAVATPTTPAP 325
 QY 321 APRTVS-----VQTLNPLAGPYGAKAGVNTLHSGVGTAAATGTTAGTGLL 365
 DB 326 EPTETIAPVVAETPEVASVAVVAETTPPVVPVVAES-----IPAVVAT----- 369

QY 366 QTSKPLVTSVANTVTVSLQPEKP-----VNSGIAVTLIS-----LPAAVT 404
 DB 370 -TTPPATIAVTDPOVTVTASAVPELPPVAPSPVSAVETPEVLDLPPVAPAEPPAVV 428
 QY 405 FGEE-----SCAATCLDSVKVYVSCFCD-----HICKPVIGTP-----VQIKL 442
 DB 429 AAEETPEPAPASAPVYTAALDIEVAPVIAAPSDAPAEADSAAPVISTPTTASVPEET 488
 QY 443 AOPGPVLSQP 452
 DB 489 APPAAVTEP 498

RESULT 10
 Q95045 PRELIMINARY; PRT; 864 AA.
 AC Q95045;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE GH08002P.
 GN CPN OR CG4795.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Tephritidae; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champagne M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuncio J., Pacleb J., Paragas V., Park S., Phuananavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;
 RU Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY058315; AAL13544.1; -
 DR Flybase: FBgn0010218; Cpn.
 DR SEQUENCE 864 AA; 84565 MW; E2B96CFB0CB9802 CRC64;

Query Match 8.9%; Score 199.5; DB 5; Length 864;
 Best Local Similarity 22.5%; Pred. No. 0.00015;
 Matches 124; Conservative 70; Mismatches 203; Indels 153; Gaps 25;

QY 1 GIVTVTV-APVSA-----PRVSGPRLPAP---QIVAKAPNTTIOFPANIQ 45
 DB 4 GIIPEVSAAPVAAVTPPSAAVAPVQVSPAAVAPAPAAPIAVTPVAPPTLASVQATAT 63
 QY 46 LPPGTVLINSNGPLMLVSPQOTVTRAETTSNTTSRPAPVAPANQVYKICVPPNS-SSQLI 104
 DB 64 IPAPAPIAASVAPVAVSAP--PVVAAPTPP-----NASVSTPPVAVAOIPVAVASAPVA 116
 QY 105 KRYAVT-----PVKKLAQIGTT--VVTIVKPPSSVOSVANPTSVVTVTGKPLNTTTLK 157
 DB 117 PVAATPTTPVAPVAPVATPPVAAAPVAPVAVTPVAVPVPVSP--VIAITPPVANTTV--- 171
 QY 158 PSSLGASSTPSPNEPNKKAENSAVQINLSPTMLE-----NKKCKNFLAMLI 204
 DB 172 PVAAPVAAVAPVAVPVLAPVAPVAVVAVETPAPPVAEIPATIPREC---AAPLI 228
 QY 205 KLASGSGSPKMGONKVLVEQLDAKIEAEETTRKLYELKSSPPHVLVPLFKKSVAL 264
 DB 229 -----PEVSVATKPLAAAEPPVAVAPPATETPVVAPAAASPVAVAPVAVETAIVA- 278
 QY 265 ROLPNSOSFIOQCVOOTSSDMVIATCTTWTTSPPVTTTSSSOSSEKSIIV-----SGAT 320
 DB 279 -----PVASAS-----TTPVAAATLTITAPET--PALAPVAVESQVAAANTVAVATPTTPAP 325
 QY 321 APRTVS-----VQTLNPLAGPYGAKAGVNTLHSGVGTAAATGTTAGTGLL 365
 DB 326 EPTETIAPVVAETPEVASVAVVAETTPPVVPVVAES-----IPAVVAT----- 369
 QY 366 QTSKPLVTSVANTVTVSLQPEKP-----VNSGIAVTLIS-----LPAAVT 404

```

Db 370 -TPVPATLAVTDPDVTASAVPELPPVIAVSPVGAVALPVPDLAPVLPVPAABEPVPAV
Qy 405 FGCT-----SGAALCPSPKPVVSEFCMD-----HICKPVIGTP-----VOIKL 442
Db 429 AEEIPERPAPASAPVTIALDIPVAPVIAAPSDAPAPASAPAIYVSTPPTTASVPETT 488
Qy 443 AOPGPVLSQP 452
Db 489 APPAVPTEP 498

RESULT 11
QYVGC8 PRELIMINARY; PRT; 864 AA.
ID QYVGC8
AC QYVGC8;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE Cpn protein.
GN Cpn OR CG4795.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Morten J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Adill J.F., Agdeyanti A., An H.-J., Andrews-Piankocch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Buttle K.C., Busman D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
RA Jaitai M., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.C.,
RA Reinelt K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weisslock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang M., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003694; AAF54754.1; -.
DR FLYbase: FBgn0010218; Cpn.
DR InterPro: IPR002965; P-rich extensn.
DR PRINTS: PR01217; PRICHEXTENSN.
SQ SEQUENCE 864 AA; 84595 MW; 5B3635A283E5C658 CRC64;

```

```

Query Match 8.9%; Score 199.5; DB 5; Length 864;
Best Local Similarity 22.5%; Pred. No. 0.00015;
Matches 124; Conservative 70; Mismatches 203; Indels 153; Gaps 25;

Qy 1 GTLVTKV-APVSAIP-----PKVSSGRLAP-----QIVAKAPNTTITGPANMQ 45
Db 4 GTIPSPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 63
Qy 46 LPFGTVLKSNSGRLMLVSPQGVTRAEVTSNITSRPVAPVAPVAPVAPVAPVAPVAPV 104
Db 64 IPAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 116
Qy 105 KKVAVT-----PVKLAQIGT--VVTVPKSSVQSVAVPTVVTVPKRLVTTTLK 157
Db 117 PPVATPPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 171
Qy 158 PSSIGASTPSNEPRKAKNSAQAQINLSPMLP-----NVKCKKFLAMLI 204
Db 172 PVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 228
Qy 205 KLAOSGSPMGQVKKLVQLDQLEADKLEAEFTRKLVLEKSSQPHLVPLKSSVAL 264
Db 229 -----PEVSVATKPLAAEBVVVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 278
Qy 265 RQLPNSQSFLOOCVOQTSSDMVATCTTVTTSPTVTTVSSSSQSEKSIIV---SGAT 320
Db 279 -----PVSAAS-----TEPPVAAATLTAPET-PALAPVAAEQVAAVNTVATPPAP 325
Qy 321 APRVVS-----VQTLNPLAGPVAKAGVTLHVSQVPIATGTTAGTGLL 365
Db 326 EPETIAPVVAETPEVAVSAVAETPPVPPVAAES-----IPAVVAT----- 369
Qy 366 QTSKPIVTSVANTVTTVSLQPEKP-----VSGTAVTLS-----LPVAVT 404
Db 370 -TPVPATLAVTDPDVTASAVPELPPVIAVSPVGAVALPVPDLAPVLPVPAABEPVPAV 428
Qy 405 FGCT-----SGAALCPSPKPVVSEFCMD-----HICKPVIGTP-----VOIKL 442
Db 429 AEEIPERPAPASAPVTIALDIPVAPVIAAPSDAPAPASAPAIYVSTPPTTASVPETT 488
Qy 443 AOPGPVLSQP 452
Db 489 APPAVPTEP 498

RESULT 12
ID 043418 PRELIMINARY; PRT; 513 AA.
AC 043418;
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Intestinal mucin (Fragment).
GN MUC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=97476275; PubMed=9334251;
RX Gum J.R., Jr., Ho J.J.L., Pratt W.S., Hicks J.W., Hill A.S.,
RA Vinal L.E., Robertson A.M., Swallow D.M., Kim Y.S.;
RT "MUC3 human intestinal mucin. Analysis of gene structure, the carboxyl
terminus, and a novel upstream repetitive region."
RL J. Biol. Chem. 272:26678-26686(1997).
DR EMBL: AF007190; AAC02268.1; -.
DR PRINTS: PR01608; BACINVASINC.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 513 AA; 52573 MW; 3BBAC2AFAAE1436 CRC64;

```

Query Match 8.8%; Score 198; DB 4; Length 513;
 Best Local Similarity 20.8%; Pred. No. 0.0001;
 Matches 110; Conservative 81; Mismatches 195; Indels 142; Gaps 22;

10 VSAPPKVSQGRPLPAQIVAKAPNT-----TTIOFP--ANLQPL-----PGTV 51
 2 MTSPSPVSS-----SITPTNTMTSMRTTYMPTANTLSPILSSILSTPVSSTE 51
 52 LIRS---NSQPL--MLVSPQOIVTRAFTSNIT--SRPVPANQIVKI-----93
 52 MISHNTNTPLSTLVLTLLTITRSPSETTYPTSPISVSDSTEIYSHITGTL 111
 94 --GVPPNSSQL--IKKVAATPVKKLAQIGTVVTPK-----PSSVQSV--AVPTS 140
 112 TATLPTSSSLPTETANMPT-----TLLTTPNTSHSTPSFTSIYSTVTS 164
 141 VVTVTPCKPLN---TVTTLKPSLSGASPSNEP-----NLK 174
 165 TPAISSASPTSGTMVSTMTMPSL--STDTPSTPTTITTPVSGTGLTATDITSTFT 223
 175 AENSAVOILSPMLENKKCNFLMLIKLACSGSQSPKGVKLYEQLDIAKIEA 234
 224 VSSSSAKSKSVIRPS--PSIONTERSLVSMTSATTPSLRPTITSTDTLSLITPST 282
 235 EEFTRKLYE-----LKSSPQPLVPEFLKSVVALAQLP-----269
 283 YSFSSSKSASAGTTHETISLSPASTNTHTTAESALATTTTSTSPTEPSTTVA 342
 270 -----NSQFIQOCVOQTSDDVATCTTVTTPSPVTVTVSSQSEKSI--IV 316
 343 TTGCGQTTPFSPSTFLETTTLPTTDFSTESLTLTAMTSPTLSTSTPTDITMSMTT 402
 317 SGATPRTVSQTLNPLAGVGAKAGVYTHSVGPRTAGTGTAGGLTOSKPLVTSVA 376
 403 SWPTATNTLSPILSSILSTSPVSTEVTTSTHT-----NTNPVSTLVTTLP--ITTR 453
 377 NTVTVSLOPEKPVSGTAVTLSP--AVTEGTSGAALCLPSKVPVVS 423
 454 STLTSEIAYSSPTSTYTESTTEITTYPTMTETSTATSLPPTSSLVS 501

RESULT 13
 ID 09QWH2 PRELIMINARY; PRT; 2045 AA.
 AC 09QWH2; 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
 DE C1 transcription factor.
 GN HCF1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kristine T.M.;
 RT cDNAs encoding the mouse homolog of the human transcription factor C1
 (HCF).
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U080821; AAD09225.1;
 DR MGD; MGT:105942; Hcfcl.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR001798; Kelch.
 DR Pfam; PF01344; Kelch; 5.
 DR SMART; SM00060; FN3; 1.
 SQ SEQUENCE 2045 AA; 210546 MW; 7AD38PCD78EBA9C CRC64;

Query Match 8.8%; Score 198; DB 11; Length 2045;
 Best Local Similarity 24.1%; Pred. No. 0.00054;
 Matches 129; Conservative 67; Mismatches 202; Indels 138; Gaps 28;

7 VAPVSAAPPKVSQGRPL---PAQIVAKAPNTTTIOFPANLQPLPGTVLIKSSGPIMLV 63
 486 VLKATGQAVTGGPPLVYMRASQ--AGKAPVTV--SLPASVRM---VVPYQAGTVIGS 539
 64 SPOOT-----VTRAETSNIT--TSRP---AVPAN-----PQVKTCTV 97
 540 NPKMSGAAALAAAAATOKIPPSAPATMSVPACTTIVKTVAVTPGTTTLPATVKAASP 599
 98 -----NSSSQLIKKVAV---TPVKLAQIGTVVTPKPSVQSVAVPTSVTVTPGKPL 150
 600 VMVSNPATRMLKTPAAQVGSVSSAANTSTRPITIVKSGTV--TVAQQAQVTVTVGCVT 658
 151 NTVYTLK--PSSL--GASPTSENEPL-----KENSAAVOINISPTMLENKKCNFL-- 200
 659 KITILVKSPTSPVSGSALISNLGKMSVQTKPVQTSAAVGAOSTGVTGIIHTKGLPA 718
 201 AMIKLKCSQSP-----EMGQVKKLVEQLDAKLEAEFTRKLVLELSSQPH 252
 719 GTILKLVTSADGKPTTITTTTQASGAGTKPTIIGI-----SSVSPS 759
 253 LVPELKKSVVALRQLPNSQFIQOCVOQTSDDVATCTTVTY-----SPVVT 302
 760 TT---KGTTLTIKTIPMSALITQAGATGVTSBGIKSPITITITTKMTSGTGAPKIIT 816
 303 -----TVSSSQSEKSLIVSGA---TAPRT-----VSQTLNPLAGPVAK- 340
 817 AVPKIATGHGQOQGVYVVLGAPGQPGTILRTVPMGCVRLVTEPVASAVPATVTVLVKG 876
 341 -AGVYTHSVGPRTAGTGTAGTGLTOSKPLVTSVA--NTVTVSLOPEKPVSGTAVT 937
 877 TTGVTTLGTYTGIVST--SLAGAGHSTASLNTPIITLCTITATLSQ---YINPAT 930
 398 LSLPVTGETSGAALCLPSKVPVSECMHICKPVIGTPEVQIKL--AOPGVLSOP 452
 931 VSAQQTTLTGAAGLTPTTINQPV-----SQGTQVLTITAPSGVGAQP 973

RESULT 14
 ID 08TE50 PRELIMINARY; PRT; 1322 AA.
 AC 08TE50;
 DT 01-JUN-2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
 DE EMSY protein.
 GN C10RF30.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hughes-Davies L.;
 RT "EMSY is amplified in breast cancer and displays a BRCA2 dependent DNA
 damage response."
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ430203; CAD22881.1;
 SQ SEQUENCE 1322 AA; 141467 MW; 7F8C95EBBA0C9F0 CRC64;

Query Match 8.8%; Score 197.5; DB 4; Length 1322;
 Best Local Similarity 23.0%; Pred. No. 0.00034;
 Matches 123; Conservative 78; Mismatches 197; Indels 137; Gaps 26;

5 TKVAPVSAAPPKVSQGRPLPAQIVAKAPNTTTIOFPAN--LQLPG--TVLIKSSGPIML 61
 149 TTSTPTSTP-----VPSGSLATYKSPRPAS---PASVNVLPVSGSTIYVAVSCSDE 197
 62 LVSPQOIVTRAFTSNITNSRPAVPANQIV---KICVPPNS---SOLIKKVA----- 108
 198 DEKPRK---RRRTNSSSSSPVLVEKPAVVPVSKTITVPSGSPKMSNIMQSIANSLLP 254
 109 -VTPVK-----KLAQIGT---VTVTPKSSVQSV-----AVPTSVTV 144

Db 255 HNSPVKITTFTKPTOTTTTQKIYITSSSTFVFNILSKSHNTAAVTKLVPTSVIAS 314
Qy 145 TPGKPLNTVTLTKPSSL-----GASSTPSNEPN-----LKAENSA 180
Db 315 TTQKP-PVITASQSLVSSSSSTSPSPITAVTAVAVSTSPVSMVTAQGST 373
Qy 181 VQINSLPMLNKKCKNFLLMLIKLACSGSPGNGVKKLVEQLDAKIEA-EEFTT 239
Db 374 SAKMASTRLPSPKSLVSAPVQIL-----AQFPKHOOSPK--QQLVQOQOQOQV 425
Qy 240 KLYELKSPQPHLVPLFKSVALROLPLNSOSFIOCCVOOTSDMIANCT-----TV 295
Db 426 PSPVSHQOQPOSPPLPKPITQIKQ-----ESGVKITIQQVOPSKILPPVATLPTS 480
Qy 296 TTSPTVTTVSSOSEKSIIVSGATAPRTVSOTLPLAGPVGAKAGVY-----345
Db 481 SNSPIMVSSNGALMTLVTTPGQATVTRPVSPISGMAAATPGAAATVKTSGSII 540
Qy 346 -----LHSG-----PAAATGTTAGTGLQTSKPLVTSVANTV---TIVSLOPEKPV 391
Db 541 TVPKSLATLGGKIISSNIVSGTTKITTIPMTSKPNVIVQKTGKGTIQLGPKMNV 600
Qy 392 S-----CTAVTSLPVTFGTSGAICLPSPKPVSEFCMDHICKPV-IGTPVQ 439
Db 601 TTLNAGEKTIQIVP-----TGAKPAILATRTPTMT---VYQPKIGSTVQ 646

RESULT 15
Q9N4S7 PRELIMINARY: PRT: 1079 AA.
AC Q9N4S7; 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Y51B11A.1 protein.
GN Y51B11A.1
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Colton M.;
RT "The sequence of C. elegans cosmid Y51B11A.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN 13
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC006797; AAF60743.1;
DR InterPro: IPR002965; P-rich_extensn.
DR PRINTS: PRO1217; PRICHEXTENS.
SQ SEQUENCE 1079 AA; 110532 MW; 8BDE3824CF80CA1 CRC64;

Query Match 8.7%; Score 195; DR 5; Length 1079;
Best Local Similarity 22.7%; Pred. No. 0.00038;
Matches 106; Conservative 50; Mismatches 206; Indels 104; Gaps 16;

Qy 5 TKVAPVSA---PKVSSGRLPAPQIVAVKAPNTTIOFPANLQLPCTVL-----IKSN 56
Db 399 TTTAPETTTSTPSSSTTP---VQTTITAPETTTSTPSSSTTPVQTTTITAPETTTST 454
Qy 57 SGPLMLVSPQOTVTRA--ETTSNTSRPAVPANQTVKICVPSNSSLIKKVAATPVK 114

Db 455 EPPSSSTTPVQTTTITAPETTS--TEPPSSSTTPVQTTTITAPETTSSTPSSSTTPVQ- 511
Qy 115 LAQIGTIVTTVPKPSVQSAVAPTSVTVTPGKPLNVTTLKPS-----LGASS 165
Db 512 -----TTTTAPETTSSTPSSSTTPVQTTTITAPETTSSTPSSSTTPVQTTTITAP 565
Qy 166 TPSNEPNLKAENSAVQINLSPMLNKKCKNFLLMLIKLACSGSPGNGVKKLVE 225
Db 566 TTSTEP---PSSSTTPVQTTTITAPETTSSTPSSSTTPVQTTTITAPETTS----- 615
Qy 226 QLLDAKIEAEFTTKLYELKSSPOPHLVPLKKSVALROLPLNSOSFIOCCVOOTSSD 285
Db 616 -----EPPSSSTTPVQTTTITAPETTSSTPSSSTTPVQTTTIT 654
Qy 286 MVIATCT-----TIVTSPVTTVSSOSEKSIIVSGATAPRTVSOTLPLA 334
Db 655 APETTSSTPSSSTTPVQTTTITAPETTSSTPSSSTTPVQTTTITAPETTSSTP- 709
Qy 335 GPVGAAGVTLHSGV-TPAATGTTAGTGLQ-----SKPLVTSVANTVTS 383
Db 710 -PSSSTTPVQTTTITAPETTSSTPSSSTTPVQTTTITAPETTSSTPSSSTTPVQTTT 768
Qy 384 LQP-----EKPVSGTAV---TSLPVTFGTSGAICLPSPKPV 421
Db 769 TAPETTSSTPSSSTTPVQTTTITAPETTSSTPSSSTTPSSV 809

Search completed: February 16, 2003, 21:59:02
Job time : 40.7465 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:40:36 : Search time 42.3161 Seconds

(Without alignments)
1738.213 Million cell updates/sec

Title: us-09-763-909-2_COPY_1_552

Perfect score: 2758

Sequence: 1 GILVTKVAPVSAAPKVVSSGP.....ASPTQKRNKIKENTSCFRDE 552

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

```
Database :
1: A_Geneseq_101002.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2758	100.0	801	19 AAW31494	Human hTAFII105 pr
2	2758	100.0	801	22 ABG09468	Novel human diagno
3	2758	100.0	852	21 AAY57279	Transcription fact
4	347	19.8	737	15 AAR56494	TATA-binding prote
5	547	19.8	737	17 AAW06084	Human TATA-binding
6	547	19.8	737	18 AAW25019	TATA-binding prote
7	521	18.9	1023	23 AAU82954	Human homologue of
8	256	9.3	921	15 AAR56487	TATA-binding prote
9	256	9.3	921	17 AAW06077	Drosophila TATA-bl
10	256	9.3	921	18 AAW25028	TATA-binding prote

11	256	9.3	921	22 ABB61528	Drosophila melanog
12	256	9.3	921	22 ABB66055	Drosophila melanog
13	221	8.0	2035	15 AAR57141	Host cell factor P
14	219	7.9	5179	22 AAM24516	C899P predicted am
15	214.5	7.8	1795	22 ABB69806	Drosophila melanog
16	212	7.7	842	22 ABB66631	Drosophila melanog
17	212	7.7	864	22 ABB71319	Drosophila melanog
18	211.5	7.7	1296	23 ABB66702	Human novel polype
19	211.5	7.7	1296	23 ABB66756	Human novel polype
20	206.5	7.5	1322	21 AAB42650	Human ORFX ORF2414
21	205.5	7.5	1920	22 AAB65656	Novel protein kina
22	205.5	7.5	2135	23 AAE21714	Human PKIN-9 prote
23	197	7.1	2972	22 AAB50363	Human SRCAP. Homo
24	197	7.1	3118	22 AAB50362	Human SRCAP. Homo
25	194.5	7.1	2781	21 AAY57453	Human transcriptio
26	194.5	7.1	2907	21 AAY57452	Human transcriptio
27	194	7.0	2971	21 AAN06725	FLO1 protein, invo
28	194	7.0	1127	22 AAB95841	Human protein sequ
29	194	7.0	1328	22 AAM78519	Human protein sequ
30	194	7.0	1331	22 AAM79503	Human protein sequ
31	194	7.0	1658	22 ABB67620	Drosophila melanog
32	194	7.0	2971	21 AAB41231	Human ORFX ORF995
33	193	7.0	862	15 AAR60563	Yeast 2.6 kb agglu
34	191.5	6.9	1714	22 ABB60186	Drosophila melanog
35	190.5	6.9	708	22 AAM79978	Human protein sequ
36	189	6.9	894	15 AAR47578	S. cerevisiae FLO1
37	189	6.9	894	15 AAR58754	Human protein sequ
38	188.5	6.8	709	22 AAM78994	Human protein sequ
39	188.5	6.8	709	22 AAB93063	Human protein sequ
40	188.5	6.8	709	22 AAB94334	Human protein sequ
41	188.5	6.8	881	22 AAM93811	Human polyepitide,
42	184.5	6.7	881	22 AAG70752	S cerevisiae apopt
43	184	6.7	1057	22 ABB65440	Drosophila melanog
44	183	6.6	557	22 AAB94078	Human protein sequ
45	182	6.6	2058	22 AAB97070	Human polypeptide

ALIGNMENTS

```
RESULT 1
AAW31494
ID AAW31494 standard; Protein: 801 AA.
XX
AC AAW31494;
XX
DT 28-APR-1998 (first entry)
XX
DE Human hTAFII105 protein.
XX
KW TATA-binding protein associated binding factor 105; human; activator;
KW hTAFII105; transcription factor; TAF10; transcriptional activation;
KW antibodies; diagnosis; therapy; biopharmaceutical industry.
XX
OS Homo sapiens.
XX
PN US5710025-A.
XX
PD 20-JAN-1998.
XX
PF 02-OCT-1996; 96US-0725012.
XX
PR 02-OCT-1996; 96US-0725012.
XX
PA (REGC ) UNIV CALIFORNIA.
XX
PI Dikstein R, Tjian R;
XX
DR WPI: 1998-109818/10.
XX
DR N-PSDB; AAV02872.
XX
PT DNA encoding human tata-binding protein associated factor - for
PT producing recombinant protein
```

XX Claim 1; Col 17-22; 12pp; English.

CC This cDNA sequence represents a human tata-binding protein associated
CC factor, htaFII105, isolated from Daudi cell nuclear extracts. Tightly
CC associated subunits (TAF's) are components of the transcription factor
CC TFIID and are thought to mediate transcriptional activation. This encoded
CC protein may be produced recombinantly from transformed host cells or
CC purified from human cells. htaFII105 specific binding agents such as
CC specific antibodies could be used for diagnosis (e.g. genetic
CC hybridisation screens for htaFII105 transcripts), therapy (e.g. gene
CC therapy to modulate htaFII105 gene expression) and in the
CC biopharmaceutical industry (e.g. as immunogens, reagents for isolating B
CC cell specific activators or other transcriptional regulators).

XX
XX
SQ Sequence 801 AA;

Query Match 100.0%; Score 2758; DB 19; Length 801;
Best Local Similarity 100.0%; Pred. No. 2.7e-191;
Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTLVTVAPVAPVAPVSSGRLPAPQIVAVKAPNTTIOFPANLQPPGVLIKNSGRL 60
DB 1 GTLVTVAPVAPVAPVSSGRLPAPQIVAVKAPNTTIOFPANLQPPGVLIKNSGRL 60
QY 61 MLVSPQOQVTRAEFTTNITSRPAVPANPQTVKICTVPSNSOLIKKVAATPVKKLAQIGT 120
DB 61 MLVSPQOQVTRAEFTTNITSRPAVPANPQTVKICTVPSNSOLIKKVAATPVKKLAQIGT 120
QY 121 TVVTVTPKSSVQSVAVPTSVTVTPGKPLNTVTLKPSLSGASTPSNEPNLKAENSAA 180
DB 121 TVVTVTPKSSVQSVAVPTSVTVTPGKPLNTVTLKPSLSGASTPSNEPNLKAENSAA 180
QY 181 VOINLSPTMLENNKKCNFLAMLIKACSGSOSPENGQVKKLYEQLDLAKIEAEFTTR 240
DB 181 VOINLSPTMLENNKKCNFLAMLIKACSGSOSPENGQVKKLYEQLDLAKIEAEFTTR 240
QY 241 LVEYELSSPOPHLVPLFKSVVALROLPLNSOSFIOQVOQTSDDVIACTTVTTSV 300
DB 241 LVEYELSSPOPHLVPLFKSVVALROLPLNSOSFIOQVOQTSDDVIACTTVTTSV 300
QY 301 VTTTSSSSOSEKSIIVSGATAPRTVSQTLNPLAGIVGAKAGVTLHSVPTAATGTTA 360
DB 301 VTTTSSSSOSEKSIIVSGATAPRTVSQTLNPLAGIVGAKAGVTLHSVPTAATGTTA 360
QY 361 GIGLLOTSRPLVTSVANVTYTSLOPEKRVYSTANTLSLPAVTBETSGAALCLPSVKR 420
DB 361 GIGLLOTSRPLVTSVANVTYTSLOPEKRVYSTANTLSLPAVTBETSGAALCLPSVKR 420
QY 421 VVSFCDHICKPVIQIPVOIKLAQPGVLSQPGIGTSSSKOLFSLFHYVOOPSGGNEK 480
DB 421 VVSFCDHICKPVIQIPVOIKLAQPGVLSQPGIGTSSSKOLFSLFHYVOOPSGGNEK 480
QY 481 OVTTTSSSTLTIOKCGQKTMPTVNTIIPTSQFPASILKOITLPGKKIISLQASPTQKNR 540
DB 481 OVTTTSSSTLTIOKCGQKTMPTVNTIIPTSQFPASILKOITLPGKKIISLQASPTQKNR 540
QY 541 IKENVTSCFRDE 552
DB 541 IKENVTSCFRDE 552

RESULT 2
ABG09468
ID ABG09468 standard; Protein: 801 AA.
XX
XX ABC09468;
AC
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #9459.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX
XX WPI: 2001-639362/73.
XX N-PsDB; AAS73655.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 20; SEQ ID No 39827; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
XX
SQ Sequence 801 AA;

Query Match 100.0%; Score 2758; DB 22; Length 801;
Best Local Similarity 100.0%; Pred. No. 2.7e-191;
Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTLVTVAPVAPVAPVSSGRLPAPQIVAVKAPNTTIOFPANLQPPGVLIKNSGRL 60
DB 1 GTLVTVAPVAPVAPVSSGRLPAPQIVAVKAPNTTIOFPANLQPPGVLIKNSGRL 60
QY 61 MLVSPQOQVTRAEFTTNITSRPAVPANPQTVKICTVPSNSOLIKKVAATPVKKLAQIGT 120
DB 61 MLVSPQOQVTRAEFTTNITSRPAVPANPQTVKICTVPSNSOLIKKVAATPVKKLAQIGT 120
QY 121 TVVTVTPKSSVQSVAVPTSVTVTPGKPLNTVTLKPSLSGASTPSNEPNLKAENSAA 180
DB 121 TVVTVTPKSSVQSVAVPTSVTVTPGKPLNTVTLKPSLSGASTPSNEPNLKAENSAA 180
QY 181 VOINLSPTMLENNKKCNFLAMLIKACSGSOSPENGQVKKLYEQLDLAKIEAEFTTR 240
DB 181 VOINLSPTMLENNKKCNFLAMLIKACSGSOSPENGQVKKLYEQLDLAKIEAEFTTR 240
QY 241 LVEYELSSPOPHLVPLFKSVVALROLPLNSOSFIOQVOQTSDDVIACTTVTTSV 300
DB 241 LVEYELSSPOPHLVPLFKSVVALROLPLNSOSFIOQVOQTSDDVIACTTVTTSV 300

Db 241 LYLKSSPOPHLVFLKSSVVALROLPLNSQSTIQOCVQOTSSDMVATCTTTVTTSVP 300
 QY 301 VTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVYTLHSGPTAATGTTA 360
 CC for treating autoimmune diseases, inflammatory processes and viral or
 bacterial infections.
 XX 301 VTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVYTLHSGPTAATGTTA 360
 SQ Sequence 852 AA;
 Query Match 100.0%; Score 2758; DB 21; Length 852;
 Best Local Similarity 100.0%; Pred. No. 3e-191;
 Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 VVSCMDHICKPVIGTPOIKLAOPGVLSPAGIPGSSSKOLFSLFHVVOQPSGGNEK 480
 Db 421 VVSCMDHICKPVIGTPOIKLAOPGVLSPAGIPGSSSKOLFSLFHVVOQPSGGNEK 480
 QY 481 QVTTISHSSTLTIOKCGKTPVNTIIPTSQPPASILKQTLTPGNKILSLQASPTQKNR 540
 Db 481 QVTTISHSSTLTIOKCGKTPVNTIIPTSQPPASILKQTLTPGNKILSLQASPTQKNR 540
 QY 541 IKENVTSQCFRDE 552
 Db 541 IKENVTSQCFRDE 552

RESULT 3
 AAR57279
 ID AAY57279 standard; Protein: 852 AA.

AC AAY57279;
 XX 06-JUN-2000 (first entry)

DE Transcription factor subunit TAFII105 polypeptide.

XX TATA box-binding protein associated factor II 105; TAFII105; cancer;
 KM transcription factor; apoptosis; cytostatic; immunosuppressive;
 KW antiinflammatory; virucide; antibacterial.

XX Homo sapiens.

OS Location/Qualifiers

FH Misc-difference 802
 FT /label= Glx

FT /note= "encoded by TGA"

FT Misc-difference 834
 FT /label= Glx

FT /note= "encoded by TAA"

XX WO200012699-A1.

XX 09-MAR-2000.

XX 25-AUG-1999; 99WO-1100464.

XX 27-AUG-1998; 98IL-0125971.

PA (YEDA) YEDA RES & DEV CO LTD.

XX Dikstein R, Yamit-hezi A;

DR WPI: 2000-256640/22.

DR N-PSDB; AA290465.

XX Polypeptide encoding TATA box binding protein associated factor II 105

PT useful for treating e.g. cancers and inducing apoptosis has a dominant

PT negative effect on the normal biological activity of the binding

PT protein -

XX Claim 7; Fig 2; 48pp; English.

CC This represents a polypeptide comprising a (modified) fragment (I) of

CC a TATA box-binding protein associated factor II 105 (TAFII105). A

CC pharmaceutical composition comprising (I) or the polynucleotide or an

CC inhibitor or antagonist of (I) is useful for treating cancers and

CC Inducing apoptosis in pathological cells. The composition is also useful

CC for treating autoimmune diseases, inflammatory processes and viral or

CC bacterial infections.

XX Sequence 852 AA;

Query Match 100.0%; Score 2758; DB 21; Length 852;

Best Local Similarity 100.0%; Pred. No. 3e-191;

Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTLVTKAVPAPVAPKSSGPRLPAPQIVAKAPNTTIQFPANLQLPFGTYLIRNSGPL 60

Db 1 GTLVTKAVPAPVAPKSSGPRLPAPQIVAKAPNTTIQFPANLQLPFGTYLIRNSGPL 60

QY 61 MLVSPQOTVTAETTSNTSPPAVPANQYKICVPPSSSOLIKKAVTVPKKLAQIGT 120

Db 61 MLVSPQOTVTAETTSNTSPPAVPANQYKICVPPSSSOLIKKAVTVPKKLAQIGT 120

QY 121 TVVTVPRKSSVQSAVPTSVVTPGKPLNTVTTLKPSISLASSSTPSNEPNLKAENSA 180

Db 121 TVVTVPRKSSVQSAVPTSVVTPGKPLNTVTTLKPSISLASSSTPSNEPNLKAENSA 180

QY 181 VOINLSPMLENVKCKKFLMLIKLACSGSOSPDMGQNVKLYBQILDAKIEAEETRK 240

Db 181 VOINLSPMLENVKCKKFLMLIKLACSGSOSPDMGQNVKLYBQILDAKIEAEETRK 240

QY 241 LVELKSSPOPHLVFLKSSVVALROLPLNSQSTIQOCVQOTSSDMVATCTTTVTTSVP 300

Db 241 LVELKSSPOPHLVFLKSSVVALROLPLNSQSTIQOCVQOTSSDMVATCTTTVTTSVP 300

QY 301 VTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVYTLHSGPTAATGTTA 360

Db 301 VTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVYTLHSGPTAATGTTA 360

QY 361 GTGLQTSKPLVTSVANTVTTSIQPEKPVSGTAIVTSLPAVTFGETSGAATCLPSVKP 420

Db 361 GTGLQTSKPLVTSVANTVTTSIQPEKPVSGTAIVTSLPAVTFGETSGAATCLPSVKP 420

QY 421 VVSCMDHICKPVIGTPOIKLAOPGVLSPAGIPGSSSKOLFSLFHVVOQPSGGNEK 480

Db 421 VVSCMDHICKPVIGTPOIKLAOPGVLSPAGIPGSSSKOLFSLFHVVOQPSGGNEK 480

QY 481 QVTTISHSSTLTIOKCGKTPVNTIIPTSQPPASILKQTLTPGNKILSLQASPTQKNR 540

Db 481 QVTTISHSSTLTIOKCGKTPVNTIIPTSQPPASILKQTLTPGNKILSLQASPTQKNR 540

QY 541 IKENVTSQCFRDE 552

Db 541 IKENVTSQCFRDE 552

RESULT 4

AAR56494

ID AAR56494 standard; Protein: 737 AA.

AC AAR56494;

XX 23-MAR-1995 (first entry)

DE TATA-binding protein-associated factor hTAFII130.

KW TATA-binding protein associated factor; hTAFII130; screening;

OS Homo sapiens.

XX WO9417087-A.

XX 04-AUG-1994.

XX 28-JAN-1994; 94WO-US01114.

XX 28-JAN-1993; 93US-0013412.

XX 28-JUN-2001; 2001MO-US20592.
 PF
 XX
 XX 29-JUN-2000; 2000US-215164P.
 PR
 XX 10-AUG-2000; 2000US-224457P.
 XX
 XX (ANAD-) ANADYS PHARM INC.
 XX
 PI Moore J, Buurman ET, Desliva T, Harris S, Komaritsky S;
 PI Mendillo M, Moore D, Mccoy M, Sanderson K, Haq T, Zhu S, Long F;
 PI Davidov E, Thompson CM;
 XX
 DR MPI: 2002-147962/19.
 N-PSDB; ABR32842.
 XX
 XX Screening candidate antifungal compound for interaction with essential
 PT protein, modulation of essential protein activity, binding to essential
 PT protein, by contacting protein with test compound and determining
 PT effects -
 XX
 PS Claim 1; Figure 79; 522pp; English.
 XX
 XX The invention describes a method of screening a candidate antifungal
 CC compound for interaction with essential proteins (EP) or for modulation
 CC of EP activity e.g. fungal gene transcription. The proteins tested in the
 CC invention include RRC34, PDB3, TRN2, NAB2, MPT1, MTR2, BOS1, PDB30, RSR2,
 CC SQT1, MTR1, TRB1, SPC98, BFR2, RNL1, GCD7, SKI6, NRP1, LCP5, NCE103,
 CC ECO1, ORC2, CNS1, YPD1, TIM10 and SRB4 from S. cerevisiae, C. albicans
 CC and human homologues. The method involves contacting a culture with one
 CC or more test compounds and determining the effects on the growth or
 CC viability of cells of the culture which preferably comprises fungal cells
 CC or yeast cells. Preferably the identified compounds interact with, or
 CC modulate (preferably inhibit) the activity of C. albicans EP. The inhibitor
 CC compounds identified by the method are useful for preventing or
 CC inhibiting fungal, particularly C. albicans growth in culture or in a
 CC mammal. The antifungal agents interact with essential fungal elements
 CC that can be used to treat fungal infection by preventing the growth and
 CC preferentially killing the fungi, but does not inhibit the biological
 CC activity of mammalian homologues. This amino acid sequence represents a
 CC target protein used to test the antifungal compounds, described in the
 CC method of the invention.
 CC
 XX
 SQ Sequence 1023 AA;
 Query Match 18.9%; Score 521; DB 23; Length 1023;
 Best Local Similarity 31.7%; Pred. No. 5.5e-29;
 Matches 168; Conservative 73; Mismatches 157; Indels 132; Gaps 22;
 QY 13 PKVSSG-----RLPAPQIVAVKAPRTTIOFPANQLPPTVLKNSGPILM 61
 DB 414 PRATSGIRATLPTVLAARLP-----PQNPNTIQ--NFQLPPGMVLVNSEGQL 464
 QY 62 LVSPQQTVR-----AETTSNTSRPAVPANPQTVKICTPVNSSQLIKKVAETPVKKL 115
 DB 465 MI-PQALAKOMQAQAHPQPTWAPRPTSPAPVQISTVQAPGPIILAR-QVTP----- 518
 QY 116 AQIGTVTVTPKPSVQGVAVPTSVVTPKGRPLTVF--TLKPSLSGASS-----TFS 168
 DB 519 ----TIIINQV--SQAQTVQPSATLQRSFGVQPLVLGGAQVTSLSLTATNAVQGTTPQ 571
 QY 169 NE-PRUKAENSAVQINLSPMLNKKKCNFLAMLIKLAGSSQSPKMGONKRLVEOD 227
 DB 572 RIVPGATTTSSAATE-----TMEVKKCKKNFLSTIKLAGSSQSTETRAVANKELVONL 625
 QY 228 LDKATIAEFTKRLVELKSSQPHLVLPLKRSVALROLPLNPSQSFIOOCVQO-----TS 283
 DB 626 LGGKIAEAEFTSLRYRELNSQPPVLPVPLKRSIPALROLTPDSAFIDQSQQPPPPS 685
 QY 284 SDMVATCTTYYTSPVVTIVSSQSEKSIIVSGATARTVSVOTLNPLAGVYGAACV 343
 DB 686 Q-----ATATLTAVALVSSSVORTAGKTAATVTSALQPPVLSL----- 722
 QY 344 VTLHSVPTAAAGTGTAGTGLQTSKPLVTSVANTVTIVSLQEPKPVSGTAVTSLPAAV 403

DB 723 -----TQPTGVGKQGPPTVLYQ-----QPKP-----GALLNPPQV 756
 QY 404 TFGETSGAALCLPSPVVPVSCMDHICKPVITGPVOIKLAQPGPVL-----SQPAGIPTGS 459
 DB 757 TLVQT-----PMVALRQPH-NRIMLTTPQGVNISEEARLILATNSELVGLTNS 804
 QY 460 SSQQLFSLEHVQO-----QPSGNEKO---VTTSHSSTLTITQKCGOK 499
 DB 805 CKDETFLQAPLQRLILEIGKKHGITELHPDVVSVYSHATQRLQNLVER 854
 RESULT 8
 AARS6487
 ID AARS6487 standard; Protein; 921 AA.
 XX
 AC AARS6487;
 XX
 DT 23-MAR-1995 (first entry)
 XX
 DE TATA-binding protein-associated factor dTAFl110.
 XX
 KW TATA-binding protein associated factor: dTAFl110; screening;
 KM diagnostic; therapeutic; gene transcription regulation.
 XX
 OS Drosophila.
 XX
 PN W09417087-A.
 XX
 PD 04-AUG-1994.
 XX
 PF 28-JAN-1994; 94MO-US01114.
 XX
 PR 28-JAN-1993; 93US-0013412.
 PR 30-JUN-1993; 93US-0087119.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Comal L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
 PI Tjian R, Wang E, Weinzierl ROJ;
 XX
 DR MPI: 1994-264019/32.
 N-PSDB; AAO70724.
 XX
 XX TATA-binding protein associated protein factors - and
 PT corresponding nucleotide sequence and deriv. antibodies, useful
 PT in screening, diagnostics and therapeutics
 XX
 PS Disclosure; Page 56-61; 180pp; English.
 XX
 CC The TATA-binding protein associated factor dTAFl110 (including
 CC specific antibodies and fusion products) are used in drug screening,
 CC diagnostics and therapeutics. They are used in the development of
 CC specific biochemical assays for screening compounds that agonise or
 CC antagonise selected transcription factors involved in regulating
 CC gene expression associated with human pathology.
 CC
 XX
 SQ Sequence 921 AA;
 Query Match 9.3%; Score 256; DB 15; Length 921;
 Best Local Similarity 22.9%; Pred. No. 7.6e-10;
 Matches 135; Conservative 86; Mismatches 193; Indels 176; Gaps 24;
 QY 25 PQIVAVKAPRTTIOFPANQLPPTVLKNSGPILM-----VSPQQTVTRAETTS 76
 DB 133 PQSPSTLSLTNTGQRA-----LVKTDNGFQLLVGTTGPPYTGITNTNSNS 184
 QY 77 NITSRAVPANPQTVKICTPVNSSQ-----LIKVAVTVPKLAQIGTVVTVTP 127
 DB 185 NITSTNHPTTQO-IRLQTVPAASMTNTATSNIIIVNSVASSGVANSQPHLQNLNQ 243
 QY 128 KPSSVQVAVPTSVVTPVTKGKPLNTVTTLKPSLSGASSPSPNEPNKAENSAVQINLSP 187

Query	Match	Similarity	Score	DB	Length	921
Best Local	Similarity	22.9%	Pred. No. 7.6e-10			
Matches	135	Conservative	86	Mismatches	193	Indels 176; Gaps
25	POIYAVKAPNTTIOFPNMLQLPCTVLKNSGPIMLT-----VSPQQTVAETTS	76				
133	POSPETILTSTLNTGQTPA-----LLAKTDNGPQLLVGTTGPPVTYQITNTNSNS	184				
77	NITSPRANPAPQYKICTVPMSSQ-----LIKVAVTVPKKLAQIGTVVTVTP	127				
185	NTTSTNHPPTTQ-TRIDTVPAAMAAMTNTTATSNITVNSVASSGVANSQPPHLLQNAQ	243				
128	KPSSVQSAVAPTSTVTVTPGKPLNVTTLKPSLIGASSPTSPNEPNLKAENSAVAQINLSP	187				
244	APQLPQIQTQITPQOQOQOVNNVSSAGGATAVASSTA-----ATT	287				
188	TWLENVK-KCKNFMLAMLIKACSGSGSPBMCQNVKKVLQDLDAKIEAEFTRKLYELK	246				
288	TOOGTKTKCKRKFANLIEL--STREPKPVKENVRILLQIELVANAYEPBFCDBRLRLN	345				
247	SSPQHLVPLFLKKSVALRQL-----LPNSQSFQ---	276				
346	ASPOCLLGLFKKSLPRLRQALYTRKELVIEGIRPPQHVGLAGLSQQLPKIQAOAIRPG	405				
277	----QCVOQTSDMVIACCTVTVTSTPVVTVTVSSQSEKSIIVSGATAPRTVS---VOT	329				
406	PSQTTITQGTQVRM--TPNALGTRPRITIGHTTIKQPPN---IRLPTAPRLVNYGIGIT	460				
330	LNPLAGPVGAKAGVTLHVSVCPTAAATGGTTAGTGILLQTSKPLVTSVANTVTTVSLQPEK	389				

QY	DB	Q1P	SLOW	PGGANIVQIR	--GPHNQLOKRTGTSVQIRATTRP	-----PNSVFPAN	-----	506
QY	390	VSGTAVTSLSPATVETGSGAICLPVSKPVVSCWDHCKIPVIGTPVQIKLADGPVL	449					
DB	507	-----KLTAVVKGOTQIKAI	-TPSLHP	-----	-PSL	530		
QY	450	SOPACIPFGSSSKQFSLFHVHQPSGSGNEKQVTTISHSSTLTIIKCGCKTWPVNTIPT	509					
DB	531	LAISGPPPTPLSLVS	-----	-TUNASIT	-----	-TLPPIPS	-LPT	564
QY	510	SOFPFASI	-----LKQITLPGN	-----KILSLQA	-----SPTOKRIKENVT	546		
DB	565	VHLPEALRAREOMONSLNHNNSHFDPAKLVEIKAPSLPHPMERINASLT	614					
RESULT 11								
ABB61528	ID	ABB61528	standard; Protein; 921 AA.					
XX	AC	ABB61528;						
XX	DT	26-MAR-2002	(first entry)					
DE	XX	Drosophila melanogaster polypeptide SEQ ID NO 11376.						
XX	XX	Drosophila; developmental biology; cell signalling; insecticide;						
KW	XX	pharmaceutical.						
XX	OS	Drosophila melanogaster.						
XX	XX	MO200171042-A2.						
XX	PD	27-SEP-2001.						
XX	PF	23-MAR-2001; 2001WO-US09231.						
PR	PR	23-MAR-2000; 2000US-191637P.						
PR	PR	11-JUL-2000; 2000US-0614150.						
XX	XX	(PEKE) PE CORP NY.						
XX	PA							
PI	PI	Venter JC, Adams M, Li PWD, Myers EW;						
XX	XX	WPI: 2001-656860/75.						
DR	DR	N-PSDB; ABL05631.						
XX	XX							
PT	PT	New isolated nucleic acid detection reagent for detecting 1000 or more						
PT	PT	genes from Drosophila and for elucidating cell signalling and cell-cell						
PT	PT	interactions -						
XX	XX							
PS	PS	Disclosure; SEQ ID NO 11376; 21pp + Sequence Listing; English.						
XX	XX							
CC	CC	The invention relates to an isolated nucleic acid detection reagent						
CC	CC	capable of detecting 1000 or more genes from Drosophila. The invention is						
CC	CC	useful in developmental biology and in elucidating cell signalling and						
CC	CC	cell-cell interactions in higher eukaryotes for the development of						
CC	CC	insecticides, therapeutics and pharmaceutical drugs. The invention						
CC	CC	discloses genomic DNA sequences (AB116176-AB130511), expressed DNA						
CC	CC	sequences (AB101840-AB116175) and the encoded proteins						
CC	CC	(ABB57737-ABB72072).						
CC	CC	The sequence data for this patent did not form part of the printed						
CC	CC	specification, but was obtained in electronic format directly from WIPO						
CC	CC	at ftp.wipo.int/pub/published_pct_sequences.						
XX	XX							
XX	XX	Sequence 921 AA;						
XX	XX	Query Match 9.3%; Score 256; DB 22; Length 921;						
XX	XX	Best local Similarity 22.9%; Pred. No. 7.6e-10;						
XX	XX	Matches 135; Conservative 86; Mismatches 193; Indels 176; Gaps 24.						
XX	XX	25 POIYAVKAPNTTITQEPANQLPQGVLIKNSGPLM-----VSPQQTVTRAEFTTS 76						

Db 133 PQSPITLSTLNTGOTPA-----LLVKTNDGFLRVGTTGTPPTVOTITNTSNNS 184
 QY 77 NITSRPAVPANPQTVKICTVPSNSSQ-----LIKVAATPVKLAQIGTVVTVTP 127
 Db 185 NITSTTNHPTTQ-IRLOTVPAAASMTNTATSNIVSVASSGANSOPHPLQLNAQ 243
 QY 128 KPSSVQSAVAVPTSVVTPGKPLNTVTLKPSLSGASSTPSNEPLKENSAAVOINLSP 187
 Db 244 APQLPQIOTIOTIPAQSQOQOVNNSAGATAVSSSTA-----ATT 287
 QY 188 TMLBNVK-KCKNFLAMLIKACSSGSPENGVNKKLYBOLLDAKIEAEFTFRKLYVEK 246
 Db 288 TQGNTEKCKRKFANLIEL-STREKPVKENVNTLIQELVNAVVEPEFCDRLERLN 345
 QY 247 SSPQHLVPLFKSVVALROL-----LPSOSFQIQ--- 276
 Db 346 ASPQCLIGFLKSLPLRQALYTKELYEGIKRPPQHVGLAGLSQQLPKIQAOIRPIG 405
 QY 277 ----QCVQOTSSDMVIACTTCTTPTTSPVYTTVSSSQSEKSIYSGATAPRTVS---VQT 329
 Db 406 -PSQTTTIGQTOVRMI--TPNALGTPRPRTIGHTTISKOPPN--IRLPTAPRLVNTGIRT 460
 QY 330 LNPAGPVGAKAGVNTLHVSPTATGTTAGTGLTOSKPLVTSVANTVTVTSIQPKP 389
 Db 461 QIP-SLOVPGQANIVQIR--GPOHAQLORTGSVOIRATRP-----PNSVPTAN----- 506
 QY 390 VVSGTAVTLSPAVTFGETSGAALCPSVKPVVSEFCMDHICKPVIGTPVQIKLAQPGVL 449
 Db 507 -----KTLAVKVGOTOIKAI-TPSLHP-----PSL 530
 QY 450 SOPAGIPTGSSSKOLFSLFHVVOOPSGGENKQVTTISHSSTLTLOKCGOKTAMPVNTIIP 509
 Db 531 AAISSGPPPTPTLSVLS-----TLNSAST-----TLPIPS-LPT 564
 QY 510 SQFPFASIT---LKOITLPGN-----KILSLQA---SPTOKNRKENV 546
 Db 565 VHLPEALRAREQOMNSLNHNSHFDKLVKIKAPSLHPHMERINASLT 614
 RESULT 12
 ID ABB66055 standard; Protein: 921 AA.
 AC ABB66055;
 XX 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 24957.
 XX Drosophila melanogaster polypeptide SEQ ID NO 24957.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 OS Drosophila melanogaster.
 XX
 PN W0200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL10158.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX Disclosure: SEQ ID NO 24957; 21pp + Sequence Listing; English.
 PS The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (AAB5737-AAB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 SQ Sequence 921 AA;
 Query Match 9.3%; Score 256; DB 22; Length 921;
 Best Local Similarity 22.9%; Pred. No. 7.6e-10;
 Matches 135; Conservative 86; Mismatches 193; Indels 176; Gaps 24;
 QY 25 PQIVAKAPNTTITQFPANLQLPCTVLKISGGLMI-----VBPQOTVTAETTS 76
 Db 133 PQSPITLSTLNTGOTPA-----LLVKTNDGFLRVGTTGTPPTVOTITNTSNNS 184
 QY 77 NITSRPAVPANPQTVKICTVPSNSSQ-----LIKVAATPVKLAQIGTVVTVTP 127
 Db 185 NITSTTNHPTTQ-IRLOTVPAAASMTNTATSNIVSVASSGANSOPHPLQLNAQ 243
 QY 128 KPSSVQSAVAVPTSVVTPGKPLNTVTLKPSLSGASSTPSNEPLKENSAAVOINLSP 187
 Db 244 APQLPQIOTIOTIPAQSQOQOVNNSAGATAVSSSTA-----ATT 287
 QY 188 TMLBNVK-KCKNFLAMLIKACSSGSPENGVNKKLYBOLLDAKIEAEFTFRKLYVEK 246
 Db 288 TQGNTEKCKRKFANLIEL-STREKPVKENVNTLIQELVNAVVEPEFCDRLERLN 345
 QY 247 SSPQHLVPLFKSVVALROL-----LPSOSFQIQ--- 276
 Db 346 ASPQCLIGFLKSLPLRQALYTKELYEGIKRPPQHVGLAGLSQQLPKIQAOIRPIG 405
 QY 277 ----QCVQOTSSDMVIACTTCTTPTTSPVYTTVSSSQSEKSIYSGATAPRTVS---VQT 329
 Db 406 -PSQTTTIGQTOVRMI--TPNALGTPRPRTIGHTTISKOPPN--IRLPTAPRLVNTGIRT 460
 QY 330 LNPAGPVGAKAGVNTLHVSPTATGTTAGTGLTOSKPLVTSVANTVTVTSIQPKP 389
 Db 461 QIP-SLOVPGQANIVQIR--GPOHAQLORTGSVOIRATRP-----PNSVPTAN----- 506
 QY 390 VVSGTAVTLSPAVTFGETSGAALCPSVKPVVSEFCMDHICKPVIGTPVQIKLAQPGVL 449
 Db 507 -----KTLAVKVGOTOIKAI-TPSLHP-----PSL 530
 QY 450 SOPAGIPTGSSSKOLFSLFHVVOOPSGGENKQVTTISHSSTLTLOKCGOKTAMPVNTIIP 509
 Db 531 AAISSGPPPTPTLSVLS-----TLNSAST-----TLPIPS-LPT 564
 QY 510 SQFPFASIT---LKOITLPGN-----KILSLQA---SPTOKNRKENV 546
 Db 565 VHLPEALRAREQOMNSLNHNSHFDKLVKIKAPSLHPHMERINASLT 614
 RESULT 13
 ID AAR57141 standard; Protein: 2035 AA.
 AC AAR57141;
 XX 19-MAR-1995 (first entry)
 DE Host cell factor protein.
 XX
 DR Herpes simplex virus; herpes virus; VP16; immediate early gene;
 XX

XX	XX	OS	Homo sapiens.	
KW	host cell factor; virus infection therapy; cellular protein			
KV	transcription.			
XX	XX	XX		
XX	XX	XX	Key	Location/Qualifiers
FT	FT	FT	Active-site	21..31
FT	FT	FT	Active-site	/note= "peptide R60"
FT	FT	FT	Active-site	168..186
FT	FT	FT	Active-site	/note= "peptide R37"
FT	FT	FT	Active-site	333..340
FT	FT	FT	Active-site	/note= "peptide R52"
FT	FT	FT	Active-site	426..449
FT	FT	FT	Active-site	/note= "peptide 362"
FT	FT	FT	Active-site	511..526
FT	FT	FT	Active-site	/note= "peptide 329"
FT	FT	FT	Active-site	578..594
FT	FT	FT	Active-site	/note= "peptide 223 first sequence"
FT	FT	FT	Active-site	594..611
FT	FT	FT	Active-site	/note= "peptide R26 1st peptide"
FT	FT	FT	Active-site	611..623
FT	FT	FT	Active-site	/note= "peptide 223 2nd sequence"
FT	FT	FT	Active-site	723..731
FT	FT	FT	Active-site	/note= "peptide 318"
FT	FT	FT	Active-site	802..813
FT	FT	FT	Active-site	/note= "peptide 299"
FT	FT	FT	Active-site	813..820
FT	FT	FT	Active-site	/note= "peptide 268"
FT	FT	FT	Active-site	836..847
FT	FT	FT	Active-site	/note= "peptide R26 2nd sequence"
FT	FT	FT	Active-site	1010..1031
FT	FT	FT	Active-site	/note= "THE TNT repeat 1"
FT	FT	FT	Active-site	1072..1093
FT	FT	FT	Active-site	/note= "THE TNT repeat 2"
FT	FT	FT	Active-site	1101..1126
FT	FT	FT	Active-site	/note= "THE TNT repeat 3"
FT	FT	FT	Active-site	1158..1183
FT	FT	FT	Active-site	/note= "THE TNT repeat 4"
FT	FT	FT	Active-site	1286..1311
FT	FT	FT	Active-site	/note= "THE TNT repeat 5"
FT	FT	FT	Active-site	1314..1339
FT	FT	FT	Active-site	/note= "THE TNT repeat 6"
FT	FT	FT	Active-site	1349..1374
FT	FT	FT	Active-site	/note= "THE TNT repeat 7"
FT	FT	FT	Active-site	1414..1439
FT	FT	FT	Active-site	/note= "THE TNT repeat 8"
FT	FT	FT	Active-site	1774..1781
FT	FT	FT	Active-site	/note= "peptide 293 2nd sequence"
FT	FT	FT	Active-site	1808..1819
FT	FT	FT	Active-site	/note= "peptide 115"
FT	FT	FT	Active-site	1819..1840
FT	FT	FT	Active-site	/note= "peptide 261 1st sequence"
FT	FT	FT	Active-site	1853..1863
FT	FT	FT	Active-site	/note= "peptide 240"
FT	FT	FT	Active-site	1901..1919
FT	FT	FT	Active-site	/note= "peptide R32"
FT	FT	FT	Active-site	1919..1930
FT	FT	FT	Active-site	/note= "peptide 261 2nd sequence"
XX	XX	XX		
XX	XX	XX	WO9413315-A.	
XX	XX	XX	23-JUN-1994.	
XX	XX	XX		
XX	XX	XX	03-DEC-1993;	93WO-US11721.
XX	XX	XX		
XX	XX	XX	04-DEC-1992;	92US-0989842.
XX	XX	XX	12-APR-1993;	93US-0046585.
XX	XX	XX		
XX	XX	XX	(COLD-) COLD SPRING HARBOR LAB.	
XX	XX	XX	(TULA-) TULARIK INC.	
XX	XX	XX		
XX	XX	XX	Herr W, Lamarco K, Wilson A;	
XX	XX	XX		

DR		WP1: 1994-234207/28.
DR	N-PsDB; AAM069229:	
PT	New Host Cell Factor polypeptide(s) and nucleic acid - are used	
PT	to develop agents for diagnosis or treatment of disease	
PT	associated with expression of a HCF-modulated gene e.g. viral	
PT	infections	
XX		
PS	Disclosure: Page 39; 71pp; English.	
CC	HCF is required for the transcription of a number of	
CC	viral genes, such as the immediate early herpes simplex virus-1	
CC	genes. Epitopes of the encoded protein can be used in	
CC	defining functional domains of HCF, identifying compounds that	
CC	associate with HCF or designing compounds capable of modifying HCF	
CC	transcription. Such agents can be used to treat viral infections.	
XX		
SQ	Sequence 2035 AA;	
	Query Match 8.0%; Score 221; DB 15; Length 2035;	
	Best Local Similarity 23.0%; Pred No. 7.9e-07;	
	Matches 131; Conservative 74; Mismatches 211; Indels 154; Gaps	27.
OY	10 VSAPKRVSSGPRLPAPQIVAVKAPNTTITIQFANILQIPGVTLIKNSNGPLMLVSPQTV 69	
Dd	537 IGSSEPMSSGMALAAATAATOKIIPSSA---PVLISVPAGTIIVKT----MAVPGTTT 588	
OY	70 TRAEFTNITSRPANPANQGVKICIVPRSSSOLIKKNAVTVYKKIAIGTV----- 122	
Dd	589 LPA--TVKVAASPVMASNPAT-----RMKLTA-----AQGTSSSATNTS 628	
OY	123 ---VTTPKPVSSQSAVAPTSVTVTPGKPLNTVTTLK-PPSL-GASTPSSNEPNL---- 173	
Dd	629 TRPIITVHKSQT-VYAQAQAVTTTVGGVTTITLVKSPIVSFGSALISNLGKMNV 687	
OY	174 --KAENSAVQINLSPTMLENKKCKNFL--AMLIKLACSGSQSP-----EMQNVK 221	
Dd	688 QTRPVQTSAVTGQASTGPVTOIIQRGPLPAGTIILKVLSADGKRPTTIITTTQASGAGTK 747	
OY	222 KLVEQLDLAKIAEETFRKLYELKSSPOPHLVPELFKRSVALRDLRPNQSQSFIOGCYQQ 281	
Dd	748 PIIIdI-----SVSEPSST--KRGTTIILIKIIPMSAIITTOAGATG 785	
OY	282 TISSDNVIATCTTYTT-----SPVYT-----TVVSSSQSEKSIIVSGA---TA 321	
Dd	786 VTSSPGIKSPITIIITTKWTSCTGAPAKIITLAVPKIATGHGOOGGYOVULKAPGQPGTI 845	
OY	322 PR-----VSVOGLNLACRYGAK--AGVVTLHSVGRPAATGTTACTGLQTS 368	
Dd	846 LRTVMGCVRLTPLYVASNAKRAVATLLVVGKGTGYTTLTGTVGST--SLAGAGCHSTS 903	
OY	369 KPLVTSVA--NTVTIVSLOPERKPVVSGTAVTLSLPAVTGEGTSGAAILCPYSKPVVSCFM 426	
Dd	904 ASLAMPRIITTLGIIVATLSSQ---VINPALIVSAAGTITLTAAGGLTPTTIIMQPV----- 954	
OY	427 DHICKPVIGTPVQIKL-AQCPVYLSQPA-GIPTGSSSKOLEBFLHNVOOPSGCNEKOVT 484	
Dd	955 -----SQPQVITLITAPEGVEAQPVHDDEVSS-----ILASP-----TT 987	
OY	485 ISHSSTLTLOKCGKTMPTVNTIIPSOPP 514	
Dd	988 EQPATVITILADSGGDVDPGVTTLVCASNPP 1017	
RESULT 14		
ID	AAM24516	
ID	AAM24516 standard; Protein; 5179 AA.	
XX	AAM24516;	
XX	12-OCT-2001 (first entry)	
DE	C899P predicted amino acid sequence.	

CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (AB5737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcf_sequences.

SQ Sequence 1795 AA;

Query Match 7 8%; Score 214.5; DB 22; Length 1795;
Best Local Similarity 20.2%; Pred. No. 2e-06;
Matches 130; Conservative 86; Mismatches 254; Indels 173; Gaps 22;

QY 9 PVSAPPKVSQPRLPAPQIVAVKAPNTTIIQFPANLQ-----PPGTVLIKNSNG 58
DB 700 PVSSTCKPPTTK-PSTRITPTTKYTTQTITTPLSSTETTSQPTTTPQPTT 758
QY 59 PLMLVSPQOTVTRAEFTSNITSRPAVPANPQTVKICTVPSNSQLIKYAVTPVKRLAQI 118
DB 759 TLTVTVPKSTTTTTEKPTSSPK-PTTQTKTSTAPNT-----KVAITQKETTP 812
QY 119 GTT-----VTVVPRSS-----VQSAVPISVTVTPGKPLNT 152
DB 813 QSTSTTIFRKTNTNPEPTSTEKPTSTPKSTTPKSTIVASSTEKTISSPKPTE 872
QY 153 VTTLKPSLSGASTPNEPNLKAENSAVAQINSPTMLENVKCKNFMAMLIKACSGSQ 212
DB 873 KSTENPTNSVKTSAFTSSTQRA-----TSTSEPTKTONITTTTPKPTLKTS 922
QY 213 SPEMGONVKLVEQLDAN--IEAEFTRLKLYELKSSPOPHLVPLKRSVALROLPN 270
DB 923 TOEATSTQKSVITTKKATFESSPLTLSTREPNTPKP-----LRTTTP 970
QY 271 SOSFIOQCVOQSSDMVATCTTVTTSPTVTTVSSSQSEKIIIVSGA--TAPR--TVS 326
DB 971 TTS-----VTAIRITTTTISSESTETSTOKPKSTPTSTRTTPKVTYI 1017
QY 327 VQTLNPLAGVAGAKGVTLHSG-----PTAATGTTA-GTGLLOTSKPL 371
DB 1018 VSTQNPPT--TTSKSTVTITTPNPSPTQRPPTTTRQPTSTASTSICITRIPTTNP 1075
QY 372 VTSVANTVTVSLQPEK-----VSGTAVTLSLPAVTEGETSGAICLPSVKPVSE 424
DB 1076 QNSTSTDLTYTRPCPDSTSDKNTACTQELQVNLLE-----LQSPKQEQP 1128
QY 425 CWDHICKPIGT-----PVQIKLAQPGVLSGPACIPTGSSSKOLFSLF 468
DB 1129 THTRHTALTGSRNTLGGQEVDPDYMDAPSSAEESGQATTAAPTMSTLAANAHLQKLF 1188
QY 469 HVV-----QPSGNEKQVTT-----ISHSFLTIQKCG 497
DB 1189 HIISTTPSRERHAPTORPSQSSQSRGVTIAOMARHNLATSKPFIASHLSLSTQOLA 1248
QY 498 ---QKTMVNTII-----PTSQFPASILKQITLPGNKIL 529
DB 1249 STOKRSIIPKTLVTNHTTKPEDESEYDSETSQYDEDENEVL 1291

Search completed: February 16, 2003, 21:54:51
Job time : 55.3161 secs


```

Db      61  MLVSPQCGVTVAETTSNTITSRPAPVANDQVTKICTVENSSSQLIKKVAVVPVKLAQIGT 120
Qy      121  TVTWTVPKRPSSVQSVNAVPTSVTVTPCKGRLNTVTYTLKRPSSLGASSTSPNEPNKAENSA 1800
Db      121  TVTWTVPKRPSSVQSVNAVPTSVTVTPCKGRLNTVTYTLKRPSSLGASSTSPNEPNKAENSA 1800
Qy      181  VOINLSPMLERNVKKCKNFLLMLIKLACSGSGSPDMGONVKKLVEQLDLDAKIEAEFTTRK 2400
Db      181  VOINLSPMLERNVKKCKNFLLMLIKLACSGSGSPDMGONVKKLVEQLDLDAKIEAEFTTRK 2400
Qy      241  LVEELKSSPQPHLYPFLKKSVALRQLLPNSQSTIQCCVQOTSDDMYIAFTCTTVTTS 3000
Db      241  LVEELKSSPQPHLYPFLKKSVALRQLLPNSQSTIQCCVQOTSDDMYIAFTCTTVTTS 3000
Qy      301  VTTVTVSSSSQSKSIIVSGATAPRTVSVQTLNPLAGPYGAKAGVYTLHSVPPTAATGTTA 3600
Db      301  VTTVTVSSSSQSKSIIVSGATAPRTVSVQTLNPLAGPYGAKAGVYTLHSVPPTAATGTTA 3600
Qy      361  GTGLLOTKRPLVTSVANTVTVTVSLQPEKPVVSGTAVTLSLPAVTFGETSGAAICLPSPVK 4200
Db      361  GTGLLOTKRPLVTSVANTVTVTVSLQPEKPVVSGTAVTLSLPAVTFGETSGAAICLPSPVK 4200
Qy      421  VVSFQWMDICRPVIGTPQIKLAPGPVLSPAPILPFGSSSKOLFSLFHVQOOPSGGNEK 4800
Db      421  VVSFQWMDICRPVIGTPQIKLAPGPVLSPAPILPFGSSSKOLFSLFHVQOOPSGGNEK 4800
Qy      481  QVTTTSHSSTLTLOKCGQKTPVNTIIPTSQFPAPASILKQITLPGNKILSLQASPTQKNR 5400
Db      481  QVTTTSHSSTLTLOKCGQKTPVNTIIPTSQFPAPASILKQITLPGNKILSLQASPTQKNR 5400
Qy      541  IKENVTSCFRDE 552
Db      541  IKENVTSCFRDE 552

RESULT 2
US-08-188-582-16
: Sequence 16, Application US/08188582
: Patent No. 5534410
: GENERAL INFORMATION:
:   APPLICANT: Tjian, Robert
:   APPLICANT: Comali, Lucio
:   APPLICANT: Dynlacht, Brian D.
:   APPLICANT: Hoey, Timothy
:   APPLICANT: Ruppert, Stegfried
:   APPLICANT: Tanese, Naoko
:   APPLICANT: Wang, Edith
:   APPLICANT: Weinzierl, Robert O. J.
:   TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
:   NUMBER OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
:   NUMBER OF SEQUENCES: 36
:   CORRESPONDENCE ADDRESS:
:   ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
:   STREET: 4 Embarcadero Center, Suite 3400
:   CITY: San Francisco
:   STATE: California
:   COUNTRY: USA
:   ZIP: 94111-4187
:   COMPUTER READABLE FORM:
:   MEDIUM TYPE: floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patentin Release #1.0, Version #1.25
:   CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/188,582
:   FILING DATE: 28-JAN-1994
:   CLASSIFICATION: 435
:   ATTORNEY/AGENT INFORMATION:
:   NAME: Osman, Richard A
:   REGISTRATION NUMBER: 36,627
:   REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
:   TELECOMMUNICATION INFORMATION:
:   TELEPHONE: (415) 781-1989

```

```

: TELEFAX: (415) 398-3249
: TELER: 910 277299
: INFORMATION FOR SEQ ID NO: 16:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 737 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-188-582-16

Query Match 19.8%; Score 547; DB 1; Length 737;
Best Local Similarity 30.2%; Pred. No. 1,4e-37;
Matches 173; Conservative 67; Mismatches 144; Indels 188; Gaps 21.

OY 13 PKVSSG-----PRLPAPQIVANVAKANTTIQPPALQOLPQGVILKSSGRLM 61
: | :|| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 68 PLATTSGRATLPTVYLAFLRLPQ-----PONPNIO-----NFLPQPGVAVRSENGQL 118
OY 62 LVSPQOATYR-----AETTSNITSRPANVPANPOTVKTCTVPSNSSQLIKKVAATPVKTL 115
: | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 119 MI-PQOALAQMOQAHAQOQTMMADRPATPTPSAPPVQISTVQAPGPIIAR-QVTP----- 172
OY 116 AQICITVVTVPKPSVSSVANPTSIVVTVPCKPLNTVY--TLKRSIAGSS-----TPS 168
: | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 173 -----TTIIKQV-----SQADITVQPSATLDRSPGVQPOLVIGAAQTASLGATAVQGTPO 225
OY 169 NE-PMLKAENSAVQINLPTMLENVKKCKNLLAMLIKLAGSGSSQSPENGVKKLVEQL 227
: | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 226 RIVPEATITSSATE-----TMEVKKCKKNFLSTLIKLAGSGKSTETAAVAKELVQNL 279
OY 228 LDKAEAEFFRKLYELKSSPQPLVPLPKSVVALROLPLNQSOTFOQCVOQ-----TS 283
: | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 280 LDKGLEADFPRLRELNSPQPLVPLFKRSLPALRLQTPDSSAFLIOSSQOCPPTPS 339
OY 284 SDMVATCTTWTTSPPVYTTVTSSSQSEKSIIVSGTAPRTVSYOTLNLPLAGPVGAKGV 343
: | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 340 Q-----ATRLTLAVVLSSVQRTAGKTAATVTSALQPPVLSL----- 376
OY 344 VTLHSVGPATAGTGTAGTGLLOTSKPLVTSVANTVTYVLSLOPEKPVVSGTAVTSLP 403
DB 377 -----TQPIQVGVGKGQPTPLVIO-----QPKP-----GALIRPQV 410
OY 404 TEGTSGAALCLPSKPVVSPFCMDHICKPVICTPVQIKLAGQGPVLSQAPAGITPSSSQ 463
: | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 411 TLTQI-----PVALRQPI-NRMLTLTPOQIDL----- 437
OY 464 LSLFHVYQPSGNEKQVTTISHSSFTLTIQKCGKTPMVNTIIPISQPPASILKQITL 523
DB 438 ----- 523
OY 524 PGNKTL---SLQASPTQKNRIKENVTSCHPDE 552
: | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 452 PGTKALSAVSAQAAAAQKNKIKEPGGSPRFD 483

RESULT 3
US-08-646-715-16
: Sequence 16, Application US/08646715
: Patent No. 3637686
: GENERAL INFORMATION:
: APPLICANT: Tjian, Robert
: APPLICANT: Comai, Lucio
: APPLICANT: Dynlacht, Brian D.
: APPLICANT: Hoey, Timothy
: APPLICANT: Ruppert, Siegfried
: APPLICANT: Tanese, Naoko
: APPLICANT: Wang, Edith
: APPLICANT: Weizsaecker, Robert O.J.
: TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS.
: TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TATS AND METHODS OF USE
: NUMBER OF SEQUENCES: 36
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT

```

```

: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/646,715
: FILING DATE: 03-MAY-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/188,582
: FILING DATE: 28-JAN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Osman, Richard A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 16:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 737 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-646-715-16

Query Match      19.8%; Score 547; DB 1; Length 737;
Best Local Similarity 30.2%; Pred. No. 1.4e-37;
Matches 173; Conservative 67; Mismatches 144; Indels 188; Gaps 21;

QY 13 PPRVSSG-----PRIPAPQIVAVKAPNTTIOFPANLOLPFGTVLIRKNSGFLM 61
Db 68 PRATTSIGRATLPFTVLAPRLPQ-----PQNTNIO---NFQLPPGMVLVRSENGLL 118
QY 62 LVBPQCVTR-----AETSNITSRAVAPNAPOTVAKICIVPNSSQILIKVAVTPKKL 115
Db 119 MI-PQQLAQMQAQAHQOPPTMAKPRATPSAPVOISTVQABTPIIAK-QVTP----- 172
QY 116 AQIGTEVTVTPKPSVQSAVPTSVTVTPGKPLNTVT--TLKPSISGASG-----TPS 168
Db 173 -----TIIKQV---SQKQTVQPSATIQRSFGVQPVQLVGGAAQTASLGTATAVQTGPQ 225
QY 169 NE-PNKAENSAVQINLSPTMLENVKCKNFMLMLIKLACSGSGSPDMGQNVKYLEOL 227
Db 226 RTVPGATTTSSAATE-----TMENVKCKNFMLSTLIKLAASSGQKQSETANVKEIVQNL 279
QY 228 LDKAIEAEFTKRLIYVLEKSSPOPHLVPLKKSVALROLIPNSOSTIOQCVQ-----TS 283
Db 280 LDKIEAEDEFTSRIRLENSPOPYLVPFLKRSPLALROLIPDSAAFIQSSQOQPPPTPS 339
QY 284 SDVVIATCTTAVTSPVTTVSSQSEKSIYSGATAPRTVSQVTLNPLAGPVGARAGV 343
Db 340 Q-----ATTALTAVLVSSVQRTACKTATVTSALQPPVLSL----- 376
QY 344 VLIHVGPRATAGTGTAGTLLQTSKPLVTSVANTVTVTSVLOPEKPVVSGTAVTSLPAV 403
Db 377 -----TQPTQVGVKQGPPLVLIQ-----QPKP-----GALIRPPOV 410
QY 404 TRGETSGAALCLPSKRVVAFCDHIKRPVIGTVOYKLAQPGVLSQAPAGIPGSSSKQ 463
Db 411 TLVQT-----PVALRQPH-NRIMLTTPQOIOL----- 437
QY 464 LFSLFHVQOPSGGNEKQVTTISHSSTLIQKCGOKTQMPVNTLIPTSOPPASILIKQTL 523
Db 438 -----NPLQPVVYVKKAVL 451

```

```

QY 524 PGNKIL---SLQASPTQKNRIKENVTSCFRDE 552
Db 452 PGTKALSAVSAQAAAAQKNKIKLEPGGSGFRDD 483

RESULT 4
US-08-188-582-2
: Sequence 2, Application US/08188582
: Patent No. 5534410
: GENERAL INFORMATION:
: APPLICANT: Tjian, Robert
: APPLICANT: Comai, Lucio
: APPLICANT: Dynlacht, Brian D.
: APPLICANT: Hoey, Timothy
: APPLICANT: Ruppert, Siegfried
: APPLICANT: Tanese, Naoko
: APPLICANT: Wang, Edith
: APPLICANT: Weinzierl, Robert O.J.
: TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
: TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
: NUMBER OF SEQUENCES: 36
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/188,582
: FILING DATE: 28-JAN-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Osman, Richard A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 921 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-188-582-2

Query Match      9.3%; Score 256; DB 1; Length 921;
Best Local Similarity 22.9%; Pred. No. 5.5e-13;
Matches 135; Conservative 86; Mismatches 193; Indels 176; Gaps 24;

QY 25 PQIYAVKAPNTTIOFPANLOLPFGTVLIRKNSGFLM-----VSPQTVRAEFTS 76
Db 133 PQSSTILSTLNTGQTPA-----LLVKTQNGFQLRVGTTGPPVTVQITNTSNN 184
QY 77 NITSRAPVAPNAPQIVKICTVFNSSQ-----LIKKVAVTPVKKLAQIGTVVTVVP 127
Db 185 NITSTNHPFTTQ--IRLOTVPAAASMTTATSNIIIVASVSSGIANSSQPPHQLQMAQ 243
QY 128 KPSSVQSAVPTSVTVTPGKPLNTVTTLLKPSISGASSTPSNEPMLKAENSAVQINLSP 187
Db 244 APQLPQIQTQIPDQSSQOQVANNVSSAGGTATAVASTTA-----ATT 287
QY 188 TMLNVK-KCKNFMLMLIKLACSGSGSPDMGQNVKYLEOLLDKAKIEAEFTKRLIYVLEK 246
Db 288 TQGGTKEKCKFLANLIEL--STREPRVKEKNVTTLIQELVANNVEDEECRDLERILN 345

```


FILING DATE: 12-APR-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Osman, Richard A
 REGISTRATION NUMBER: 36,627
 REFERENCE/DOCKET NUMBER: A-57503-1/RAO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELEX: 910 277299
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2035 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-046-585-5

Query Match 8.0%; Score 221; DB 1; Length 2035;

Best Local Similarity 23.0%; Pred. No. 1.6e-09;
Matches 131; Conservative 74; Mismatches 211; Indels 154; Gaps 27;

QY 10 VSAPPKVSSGRLPAQIVAVKAPNTTITQFPANLQLPFGTVLIKNSGMLVSPQQT 69
 DB 537 IGSSPQSGMAALAAATAATOKIPSSA---PYLVSPAGTTIVKT-----MAVTPGTTT 588
 QY 70 TRAEITNITSRPVAVPOTVKTICVNVSSSOLIKKVAATPVKKLAQIGTTV----- 122
 DB 589 LPA--TVKVASPVVAVSNPAT-----RMLKTRA-----AOGVSASATNTS 628
 QY 123 ---VTVPKPSVAVPNTSVTVTPGKPLNTVTLK-PSSL-GASSTPSNEPNTL---- 173
 DB 629 TRPIITVHKSGTV-TVAQQAQVTVTVGVTKITLVKSPISVPGSALISNLGKVM 687
 QY 174 --KAENSAVQINISPTMLENVKCKNFL--AMLIKACSGSOSP-----EMGQNVK 221
 DB 688 QTRPVQTSVAVTGAQSTGSPVQIIOTKGRLPAGTILKLVTSADGKPTIITTTQASGAGTK 747
 QY 222 KLVEQLDAKIEAEFEFRKLYVELKSSPQHLVFLKSSVALROLPLNSQSFIOQCQVQ 281
 DB 748 PTILGI-----SSVSPSTT---KPGTTTITKTIPMSAIIITQACATG 785
 QY 282 TSSDMVATCTTCTT-----SPVVT-----TVSSSSEKSIIVSGA-----TA 321
 DB 786 VTSSPGIKSPIITITTKVMTSGTGAPAKIITAVPKIATGHGQGVTVVLKGAAGQPGTI 845
 QY 322 PRT-----VSQVTLNPLAGPVGAK--AGVTLHSVGPPTAATGTTAGTGLQTS 368
 DB 846 LRTVPMGQVRLVTPVTSVAVKPAVTTLVVKGTTGTVTGIVST--SLAGAGHSTS 903
 QY 369 KPLVTSVA--NTVTVSLOPEKPVSGTAVTSLPAPVFGETSAGALCLPSVKPVSVFCW 426
 DB 904 ASLATPTITLTGTTATLSSQ---VINPTAITVSAQTTILTAAGGLTPTTMOVP----- 954
 QY 427 DHICKPVIGTPVQIKL-AQGPVLSQPA-GIPTSSSKOLFSLFHVVOQPSGNEKQVTT 484
 DB 955 -----SQPTQVTLTATASGVAEQPVHDLPS-----ILASP-----TT 987
 QY 485 ISHSSTLTQKCGOKTMPVNTIITPSOFP 514
 DB 988 EOPTATVTLADSGQDVQPGTIVTLVCSNPP 1017

RESULT 7
 US-08-393-703-5
 ; Sequence 5, Application US/08393703
 ; Patent No. 558239
 ; GENERAL INFORMATION:
 ; APPLICANT: Lamarco, Kelly
 ; APPLICANT: Wilson, Angus
 ; APPLICANT: Heit, Winship
 ; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:

TITLE OF INVENTION: HOST CELL FACTOR
 NUMBER OF SEQUENCES: 15
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: CA
 COUNTRY: USA
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/393,703
 FILING DATE: 24-FEB-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Osman, Richard A
 REGISTRATION NUMBER: 36,627
 REFERENCE/DOCKET NUMBER: A-57503-2/RAO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELEX: 910 277299
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2035 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-393-703-5

Query Match 8.0%; Score 221; DB 1; Length 2035;

Best Local Similarity 23.0%; Pred. No. 1.6e-09;
Matches 131; Conservative 74; Mismatches 211; Indels 154; Gaps 27;

QY 10 VSAPPKVSSGRLPAQIVAVKAPNTTITQFPANLQLPFGTVLIKNSGMLVSPQQT 69
 DB 537 IGSSPQSGMAALAAATAATOKIPSSA---PYLVSPAGTTIVKT-----MAVTPGTTT 588
 QY 70 TRAEITNITSRPVAVPOTVKTICVNVSSSOLIKKVAATPVKKLAQIGTTV----- 122
 DB 589 LPA--TVKVASPVVAVSNPAT-----RMLKTRA-----AOGVSASATNTS 628
 QY 123 ---VTVPKPSVAVPNTSVTVTPGKPLNTVTLK-PSSL-GASSTPSNEPNTL---- 173
 DB 629 TRPIITVHKSGTV-TVAQQAQVTVTVGVTKITLVKSPISVPGSALISNLGKVM 687
 QY 174 --KAENSAVQINISPTMLENVKCKNFL--AMLIKACSGSOSP-----EMGQNVK 221
 DB 688 QTRPVQTSVAVTGAQSTGSPVQIIOTKGRLPAGTILKLVTSADGKPTIITTTQASGAGTK 747
 QY 222 KLVEQLDAKIEAEFEFRKLYVELKSSPQHLVFLKSSVALROLPLNSQSFIOQCQVQ 281
 DB 748 PTILGI-----SSVSPSTT---KPGTTTITKTIPMSAIIITQACATG 785
 QY 282 TSSDMVATCTTCTT-----SPVVT-----TVSSSSEKSIIVSGA-----TA 321
 DB 786 VTSSPGIKSPIITITTKVMTSGTGAPAKIITAVPKIATGHGQGVTVVLKGAAGQPGTI 845
 QY 322 PRT-----VSQVTLNPLAGPVGAK--AGVTLHSVGPPTAATGTTAGTGLQTS 368
 DB 846 LRTVPMGQVRLVTPVTSVAVKPAVTTLVVKGTTGTVTGIVST--SLAGAGHSTS 903
 QY 369 KPLVTSVA--NTVTVSLOPEKPVSGTAVTSLPAPVFGETSAGALCLPSVKPVSVFCW 426
 DB 904 ASLATPTITLTGTTATLSSQ---VINPTAITVSAQTTILTAAGGLTPTTMOVP----- 954
 QY 427 DHICKPVIGTPVQIKL-AQGPVLSQPA-GIPTSSSKOLFSLFHVVOQPSGNEKQVTT 484

Db 955 -----SQPQVTLITAPSGVEAQPVDLPVS-----ILASP-----TT 987

QY 485 ISHSTLTTCGCGQKTMPTVTITPTISQFPP 514

Db 988 EQPATVTLTADSGQGDVQPGTIVLVCNNP 1017

RESULT 8

PCT-US93-11721-5

Sequence 5, Application PC/TUS9311721

GENERAL INFORMATION:

APPLICANT: Lamaco, Kelly

APPLICANT: Wilson, Angus

APPLICANT: Hart, Winship

TITLE OF INVENTION: A NOVEL EUKARYOTIC TRANSCRIPTION PROTEIN:

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESS: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/11721

FILING DATE: 03-DEC-1993

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Osman, Richard A

REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: FP-57503-1/RAO

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 396-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 2035 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

PCT-US93-11721-5

Query Match 8.0%; Score 221; DB 5; Length 2035;

Best Local Similarity 23.0%; Pred. No. 1.6e-09;

Matches 131; Conservative 74; Mismatches 211; Indels 154; Gaps 27;

QY 10 VSAPKPVSSGPRLPAPQIYAVKAPNTTITQFPANLQIPGTIVLIRKNSGRLMVSQOQY 69

Db 537 IGSQSMGSMALAAAAATOKIPSSA---PVLSPVPGTITVKT---NAVTGCTT 588

QY 70 TRAEFTSNITSRPAPVAPNDQVKTCTVPSNSQOLIKKVAATPVKKLAQIGTIV-----122

Db 589 LPA--TVKVAASSPVMSVSNPAT-----RMLKTA-----AQGTSVSASATMS 628

QY 123 ---VTVTPKPSVQSAVATSVYTVTPGKPLNTVTLK-PSL-GASSTPSNEPNL-----173

Db 639 TRPITTVHSGTV-TVAAQAAQVTVTVGVTTKTTLVKSPIVSIGSALISNLGKVMYV 687

QY 174 --KAENSAVQINTLSPITMLNKKCNFL--AMLIKLAGSGOSP-----EMGQNVK 221

Db 688 QKRPVQTSVAVTQASGTPVTOIITKGPLPAGTITLKVMSADCKPTTITTTTQASGAGTK 747

QY 222 KLVEQLDAKIAEETRKLYVELKSSPOPHLVPLKKSVALROLPLPSQSEFIQOCVQO 281

Db 748 PTLIGI-----SSVSPSTT---KPGTTTIIKIIPSAITTOAGANG 785

QY 282 TSDMVAIATCTTVT-----SPVVT-----TVSSSGSEKSIIVSGA-----TA 321

Db 786 VTSPPGKISPTITITTKVMTSGTGAPAKITPAKIAITGCGQGVTVLKLGAQCQGTI 845

QY 322 PRT-----VSQVTLNPLAGPVGAK--AGVTLHSGVPAATGTTAGTGLQTS 368

Db 846 LRTVPMGVRVLPVTVSAKPAVTVTLVKGTTGTVTGTVST--SLAGAGHSTS 903

QY 369 KPLVTSVA--NTVTVSLQPEKPVSGTAATLSLPAVTFGETSGAICLPSVFPVSCW 426

Db 904 ASLATPITLTGTTATLSSQ---VINFTATVSAQVTLTAAGITPTITMQPV-----954

QY 427 DHICKPVITPVOIKL-AQGPVLSQPA-GIPGSSSKQLSLFHVYQDPSGGENQVTT 484

Db 955 -----SQPQVTLITAPSGVEAQPVDLPVS-----ILASP-----TT 987

QY 485 ISHSTLTTCGCGQKTMPTVTITPTISQFPP 514

Db 988 EQPATVTLTADSGQGDVQPGTIVLVCNNP 1017

RESULT 9

US-09-579-181-2

Sequence 2, Application US/09579181

Patent No. 6365372

GENERAL INFORMATION:

APPLICANT: Chivlia, John

APPLICANT: Yaciuk, Peter

TITLE OF INVENTION: SNE2 Related CBP Activator Protein (SRCAP)

FILE REFERENCE: 16153-4247

CURRENT APPLICATION NUMBER: US/09/579,181

CURRENT FILING DATE: 2000-05-25

PRIOR APPLICATION NUMBER: 60/136,620

PRIOR FILING DATE: 1999-05-27

NUMBER OF SEQ ID NOS: 17

SOFTWARE: Patent In Ver. 2.0

SEQ ID NO 2

LENGTH: 2972

TYPE: PRT

ORGANISM: Human

US-09-579-181-2

Query Match 7.1%; Score 197; DB 4; Length 2972;

Best Local Similarity 23.7%; Pred. No. 2.9e-07;

Matches 140; Conservative 75; Mismatches 240; Indels 136; Gaps 32;

QY 2 TLVTKVAPVSAAPPKVSQ-----PRLPAPQIYAVKAPNTTITQFPANLQ-- 46

Db 1063 TVPPLAPAPRPP--SSGLPAVLNPRPTLTGGRUPPTLTGTAPAPMPTPLVRLPLKVH 1120

QY 47 --PPGTVLIRKNSGRLMVSPOQTVTRAETTSNITSRPAPVAPNDQVKTCTVPSNSQOLI 104

Db 1121 SPSPVSAASAPGAAPLTISSPLHP-----SSLPGPASSPM-----PIRSSPLAS 1166

QY 105 KKAIVTPVKKLAQIGTIVTVTPKPSVQSAVAPTSV-VTVTPGKP-LNTVT-TLKPS 160

Db 1167 PVSTSVSPVLSLSPISVPTLTPAPASA-PLTIPISAPVLSASGAPALLTSVPPPLPV 1225

QY 161 IGASSTPSNEPNLKAENSAVQINTL--SPITMLNKKCNFLMLIKLAGSGOSPSEMGO 218

Db 1226 PAAFGPSPVLPQSGASPSASALITGLATAPSLSSQTPGPHL-----LAFTSHVGLNS 1280

QY 219 NVKRLVEQ-LDAKIAEETRKLYVELKSSPOPH-----LVPLKKSVALROLPLNS 271

Db 1281 TVAPACSPVLPVPSALASPP-----PSAPRPAPQASLAPASASGALATPLAPMA 1332

QY 272 QSFITQCVQOQTSMDVAIATCTTVTTSPTVTVSSQSEKSIIVSGAT-----320

Db 1333 AP--QVAILAPSPAPPLAP-LPVLAPSPGAPVPLASQTPVPMAPISSTFGTSLASAPV 1389

QY 321 -APRTV-----SVQTL--NPLAGVGAAGVTVLHSGVPTAA--TGCT-----TAITGL 365

Db 1390 PATPVLABSSTQTMPLPVPSPSPASTQTL-ALAPALAPLGGSSPQSTLSLGTGNP 1448
Qy 366 QNSKPLVTSVANTVTTSLOPEKPVYSGTAVTSL-PAVTEGETSGAIC-LPSVKPVYS 423
Db 1449 QGFEP-----TQTLSTLPASSLVPTPAQTLSTLAPGPPGLGTQTLSTLAPAPPLAP--- 1497
Qy 424 FCMDDHICKPVITGV-----QIKLAQGPVUS---QPGCITP-GSSSKQLSFLH 469
Db 1498 -----ASPGAPAPAHTLTLAPASSSASLAPASVQTLTLSPAPVPTLGPMAAQTLALAP 1551
Qy 470 V-VOQPSGNEKQVTTISHSSTLTIQCGOKTAPVNTI--IPASOFPASI 517
Db 1552 ASTQSPA-----SQASSLVASASGAPPLPVTWVKPLVSKDEPDTL 1592

RESULT 10
US-09-579-181-1
; Sequence 1, Application US/09579181
; Patent No. 635372
; GENERAL INFORMATION:
; APPLICANT: Chiviva, John
; APPLICANT: Yaclick, Peter
; TITLE OF INVENTION: SMT2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247 US/09/579,181
; CURRENT APPLICATION NUMBER: US/09/579,181
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,620
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 1
; LENGTH: 3118
; TYPE: PRT
; ORGANISM: Human
US-09-579-181-1

Query Match 7.1%; Score 197; DB 4; Length 3118;
Best Local Similarity 23.7%; Pred. No. 3, 2e-07;
Matches 140; Conservative 75; Mismatches 240; Indels 136; Gaps 32;

Qy 2 TLVTKPAVSPAPKVS-----PLTPAQIVAVKAPNTTITQFPAQTL-- 46
Db 1209 TTPVPLAPAPRP--SSGLPAVLNPPPTLTGRLPTTGTATAPMTPTLVNPLKLVH 1266
Qy 47 --PQGTVLKSNQPLMLVSPQQTVAETTSNITSRPANPQVTKICTVNSSQLI 104
Db 1267 SPSEVASASAPGAPLTISSPLHP-----SSLPGASSPM-----PINSSPLAS 1312
Qy 105 KKAAYTPVKKLAQIGTTVTTPKSSVQSAVPTSV-VTVTPGKP-LMTVT-TLKPS 160
Db 1313 PVSSTVSPSSSLPISVPTTLPAPASA-PLTIPISAPLTVSASGPAALTSVPPLAPV 1371
Qy 161 LGASSPSPNEPNKKAENSAVOINL--SPTMLENVKKCKKFLMLIKLACSSQSPBMQ 218
Db 1372 PAAPGPPSIQPSASPSASALTLGLATAPLSSSQTPGHPL-----LAPTSHVPLNS 1426
Qy 219 NVKKLVEQ-LLDKIAEEETTRKLYELKSSQPH-----LVFLKKSVALROLPLNS 271
Db 1427 TVAPACSPVLVPASALASP-----PSAPAPAPQASLAPASASALATPLAPMA 1478
Qy 272 QSFITQCVQOOTSDMVIACTTTVTTSVVTTVSSSSQSEKSIIVSGAT----- 320
Db 1479 AP--QTAALAPSPAPPLAP-LPVLAPSPGAPVLAASQTPVPVMASSPSTGSLASAPV 1535
Qy 321 -APRTV-----SVQTL--NPLAGPVGAKAGVTVTHSVGPTAA--TGCT-----TAGTGL 365
Db 1536 PATTPVPLASSITQTMPLPAPVPSPLPSPASTQTL-ALAPALAPLGGSSPQSTLSLGTGNP 1594
Qy 366 QNSKPLVTSVANTVTTSLOPEKPVYSGTAVTSL-PAVTEGETSGAIC-LPSVKPVYS 423
Db 1595 QGFEP-----TQTLSTLPASSLVPTPAQTLSTLAPGPPGLGTQTLSTLAPAPPLAP--- 1643
Qy 424 FCMDDHICKPVITGV-----QIKLAQGPVUS---QPGCITP-GSSSKQLSFLH 469

Db 1644 -----ASPGAPAPAHTLTLAPASSSASLAPASVQTLTLSPAPVPTLGPMAAQTLALAP 1697
Qy 470 V-VOQPSGNEKQVTTISHSSTLTIQCGOKTAPVNTI--IPASOFPASI 517
Db 1698 ASTQSPA-----SQASSLVASASGAPPLPVTWVKPLVSKDEPDTL 1738

RESULT 11
US-08-325-267A-4
; Sequence 4, Application US/08325267A
; Patent No. 5585271
; GENERAL INFORMATION:
; APPLICANT: MATARI, JUNJI
; APPLICANT: TAKATA, YOSHIRO
; APPLICANT: OGAWA, MASAHITO
; APPLICANT: PENTTILA, MERJA
; APPLICANT: ONNELA, MAIJA-LEENA
; TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,267A
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP PCT/JP94/00290
; FILING DATE: 24-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 38871/1993
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2589-023-0XPCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 862 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-325-267A-4

Query Match 7.0%; Score 193; DB 1; Length 862;
Best Local Similarity 20.1%; Pred. No. 1e-07;
Matches 120; Conservative 86; Mismatches 216; Indels 174; Gaps 24;

Qy 11 SAPPKVSGRLPAPQIVAKAPNTTITQF---PANLQLPQGTVLKSNQPL----- 60
Db 199 SLPPNIEGTYVMAGYGYPPKVVYSNAVSWGTLPISVTLPLDGTTVSDDEGYVSEDDDL 258
Qy 61 -----MLVSPQQTVAET-----TSNISRRAVNPANPQVTKICTVNSSQLI 99
Db 259 SQSNCTVPPDSNNAVSTTTTTEPWTGTFTSTEMTFTYGTGNGVFPD-ETVIVIRTPT 317
Qy 100 SSQILKRVAVTPVKKLAQIGTTVTTPKPSVQSAVPTSVTVTPGKPLN-TVTTLK- 157

APPLICATION NUMBER: US/08/971,692
 FILING DATE:
 CLASSIFICATION: 435
 INFORMATION FOR SEQ ID NO: 15:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 894 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-971-692-15

Query Match

6.9%; Score 189; DB 3; Length 894;
 Best Local Similarity 20.7%; Pred. No. 2,3e-07;
 Matches 118; Conservative 84; Mismatches 217; Indels 152; Gaps 24;

QY 22 LPAPQIVAKKAPNTTIOFPANLQLPQTVLIKNSGPIMLVSPQO-----TVTRAE 73
 DB 265 VPDSNNAVSTTTT-----EPWTGFTSTSTMTVTGTNGVPTDEVIIVIRP 315
 QY 74 TTSNITSRPVP-----ANPQTVKICVPPNSSQILKKVAV--TPVKLAQIGTIVTV 126
 DB 316 TSEGLISTTEPMTGFTSTSTMTVTGTNGVPTDEVIIVIRPSE---GLISTTE 371
 QY 127 PKSSVQAVAVPTSVVTPKPLN-TVTLK-PSSLGASTPSNEP-----NLKAENS 178
 DB 372 PMTGFTSTMTVTGTNGVPTDEVIIVIRPSEGLVTT-TPEPTGFTSTSTMS 430
 QY 179 AAVQINISPTMLENVKCKNFAMLIKACSGSPKMGQVKKVLEQLDAKTEAEFT 238
 DB 431 TVTGTNGLPTD-ETVIIVKT-PTTAISSLSSSSSGQITSI----- 470
 QY 239 RKLVEKSSFPQHLVPLKKSVALRQLPNSQSFIOQVQOQSSDMVATCTTV-TT 297
 DB 471 -----TSSRPITTFEYPS-----NGTSVI-----SSSVISSVTSLSPTS 505
 QY 298 SPVYTTVSS-----SQSEKTIIV-----SGATRPVSVQVNLNLAGPVGKA 341
 DB 506 SPVSSSVISSTTST 558
 QY 342 GVVTLHSGPAAATAGTGTAGLQTSKPLVTSVANTVTVTSLOPEKPVSGTAVTSLP 401
 DB 559 FISESSKSPYSS-----SLPLVTSATSETMS-----SLP 592
 QY 402 AVTFGETSGAICLPKPVVSCWDHIC---KPVIGTPVOIKLAGPVLSPAGIPT 457
 DB 593 PATTKISSEQTLV-----TVTSCSHVCTESIPAIYSTATVIV-----SGVTT 637
 QY 458 GSSSKQLEFLHVVQPSGNEKQVTTISHSSITLQKCGKTPVNTIIFTSQPPASI 517
 DB 638 EYTWCPISSTETTKQKGTTEQTEETKQTVVTTISSC-ESDVCSTKASPAIVSTSTAT 696
 QY 518 LKQITLPGNKILSLQASPTOKNRKIKENVTS 548
 DB 697 INGVTEYTWCPISTESTESRQQTLLVTVTSC 727

RESULT 14
 US-08-325-267A-2
 Sequence 2, Application US/08325267A
 Patent No. 5585271

GENERAL INFORMATION:
 APPLICANT: MATSURI, JUNJI
 APPLICANT: TAKATA, YOSHIHIRO
 APPLICANT: OGAWA, MASAHIRO
 APPLICANT: PENTTILA, MERJA
 APPLICANT: ONNELA, MAIJA-LEENA
 APPLICANT: KERANEN, SIRKA
 TITLE OF INVENTION: YEAST AGGLOUTINATION GENES AND YEAST
 TITLE OF INVENTION: CONTAINING THEM
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: OBION, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
 STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400

CITY: ARLINGTON
 STATE: VA
 COUNTRY: USA
 ZIP: 22202
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/325,267A
 FILING DATE: 18-NOV-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP PCT/JP94/00290
 FILING DATE: 24-FEB-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 38871/1993
 FILING DATE: 26-FEB-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: OBION, NORMAN F
 REGISTRATION NUMBER: 24,618
 REFERENCE/DOCKET NUMBER: 2589-023-0XPCF
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-413-3000
 TELEFAX: 703-413-2220
 TELEX: 248855 OPAT UR
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1537 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-325-267A-2

Query Match 6.5%; Score 179.5; DB 1; Length 1537;
 Best Local Similarity 20.5%; Pred. No. 3.2e-06;
 Matches 109; Conservative 82; Mismatches 195; Indels 147; Gaps 22;

QY 50 TVLI---KNSGPIMLVSPQOIVRAETSNITSNPANPAP--QTVKICVPPNSSQIL 104
 DB 938 TVIIVIRPTEGLISTTEPMTGFTSTSTMTVTGTNGVPTDEVIIVIRPSE--- 993
 QY 105 KKVATPVKLAQIGTIVTVTPKSSVQAVAVPTSVTVTPKPLN-TVTLK-PSSLG 162
 DB 994 -----GLISTTEPMTGFTSTSTMTVTGTNGVPTDEVIIVIRPSE 1039
 QY 163 ASSTPSNEP-----NLKAENSAVQINLSPTMLENVKCKNFAMLIKACSGSPKMG 216
 DB 1040 LVTT-TPEPTGFTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1096
 QY 217 GQNVKQVLEQLDAKTEAEFTKRIYVELKSSPQHLVPLKKSVALRQLPNSQSTIQ 276
 DB 1097 TTSI-----TSSRPITTFEYPS-----NGTSVI- 1119
 QY 277 QCVQOTSSDMVATCTTV-TTSPVYTTVSS-----SQSEKTIIV-----SGA 319
 DB 1120 -----SSSVISSVTSLSPTSSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1173
 QY 320 TAPRIVSVQTLNPLAGPVGAKAGVTVLHSGVPTAATGTTAGTGLQTSKPLVTSVANTV 379
 DB 1174 SESETSS-----AGSVSSSSSFISESSKSPYSS-----SLPLVTSATTSQ 1215
 QY 380 TVTSLOPEKPVVSGAVATLSLPATVFGTSGAICLPKPVVSCWDHIC---KPVIG 435
 DB 1216 ETAS-----SLPPATTTTSEQTIV-----TVTSCSHVCTESIPAIIV 1255
 QY 436 TPVOIKLAQPPVLSQPAIGFTGSSKQLEFLHVVQPSGNEKQVTTISHSSITLQK 495
 DB 1256 STAVTV-----SGVTEYTWCPISSTETTKQKGTTEQTEETKQTVVTTISS 1305
 QY 496 CGQKTPVNTIIFTSQPPASILKQITLPGNKILSLQASPTOKNRKIKENVTS 548

Db 1306 C-ESDVCSTKSPAIYSTATNGVTTEYTWCPISTESROOTLIVVTSC 1357

Search completed: February 16, 2003, 22:02:41
job time : 29.1842 secs

```
RESULT 15
US-08-700-651-5
; Sequence 5, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAs AND RNAs
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1721
; TYPE: PRT
; ORGANISM: Cryptosporidium parvum
US-08-700-651-5

Query Match          6.3%; Score 174.5; DB 3; Length 1721;
Best local Similarity 20.9%; Pred. No. 1e-05;
Matches 118; Conservative 65; Mismatches 237; Indels 145; Gaps 23;

QY 35 TTTTIOEPANIQLPPTGTVLTKNSGSLMLVSPQGTVTRAETTSNITSRAVAVANPQTVKIC 94
DB 286 TTTTITTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 345
QY 95 TVNSSSQLKKAAVTPVKLAQIGTVTVTVKPSVQSAVPTSVYTPGKPLNTVT 154
DB 346 TTTTITTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 398
QY 155 TLKPSISLIGASSTPS-----NEPLKAENSA-----VOINLSPTMLEN-- 192
DB 399 TTTATTTTTTSETSVIKPDCEMCLEKNGECEAGATVGVIGKDGRIENGMATMIPND 458
QY 193 -----VKKCNFLAMLIKLACSGSQSPGCONVAKKIVEQLDAKIE--AEEFTRKLY 242
DB 459 THVRFREKYADVONTISVRCRKAGKLEFP-----DRSLDEFTIPVACHNSCSII 508
QY 243 VELKSSPOPLVPLFKRSVALRQLPNSQSF-----IQQCV-----QOTSSDMYIA 289
DB 509 VGVSGDGKIHVSPYGSKDVSLISAPIOPSELPNEVYCDCTAKYGAHSGYOTSADFTVT 568
QY 290 TC-----TTVTTSFVTVTVSSOSEKSIIVSGATAPRTVSVOQLNPLAGFVGAAGVTL 346
DB 569 TTKKPTTTTGGAGGPTTTTTSKSPK--TTTTTKATTTTTTLNP-----IITF 616
QY 347 HSYGPTAAGGTAGTGLQTSKPLVTSVANVTVVSLOPEKPV-----SGTAVTSL 400
DB 617 TTQKPTTTT--TTKVPG-----KP--PIATTTTTL-----KPIVTTTTTKATTTTTT 661
QY 401 PAVTFETSGAALCLPVSVPVSVFCMDHICKPIY-----TPVOIKLAQPGVLSQPRAGIP 456
DB 662 PTTT--TTTKRDEMTTTTTTPPL-----PDIGIEITPIPIE----- 694
QY 457 TGSSSKQLSFLFHVVOOPSGGNEKQVTTISHSSTLTIQKCGKTMPTVNTIITPSOPPPAS 516
DB 695 -----KMLDKYTMITDYNSG-----LLDSNDPITPSQAGQ-----IADTSNLEPVQ 738
QY 517 LKQITLPGNKILSLQASPTQKNRI 541
DB 739 THKSTGLPIDPMVGLPDPKSGNLV 763
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:56:34 ; Search time 10.8745 Seconds

(without alignments)
1296.883 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_552

Perfect score: 2758
Sequence: 1 GTLVTKVAPVSAAPKVSSGP.....ASPTQKRIKENTSCFRDE 552

Scoring table: BLOSUM62
Gapop 10.0 , Gaepxt 0.5

Searched: 140259 seqs, 25548876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA.*
1: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubppa/PCr_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/2/pubppa/PCrUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/2/pubppa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/2/pubppa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	521	18.9	1023	9 US-09-893-519A-14	Sequence 14, Appl
2	220	8.0	1367	10 US-09-801-368-108	Sequence 108, App
3	219	7.9	5179	9 US-10-025-380-1068	Sequence 1068, Ap
4	219	7.9	5179	10 US-09-922-217-1068	Sequence 1068, Ap
5	219	7.9	5179	10 US-09-833-263-1068	Sequence 1068, Ap
6	195	7.1	1322	10 US-09-801-368-114	Sequence 114, App
7	179.5	6.5	1537	10 US-09-801-368-104	Sequence 104, App
8	177.5	6.4	2665	10 US-09-864-761-34248	Sequence 34248, A
9	176.5	6.4	1601	10 US-09-862-027-40	Sequence 40, Appl
10	174	6.3	1075	10 US-09-862-027-40	Sequence 40, Appl
11	172	6.2	1169	10 US-09-801-368-110	Sequence 110, App
12	159.5	5.8	1056	9 US-10-161-510-10	Sequence 10, Appl
13	158.5	5.7	2597	10 US-09-905-129-2	Sequence 2, Appl1
14	158.5	5.7	2597	10 US-09-905-129-10	Sequence 10, Appl
15	158.5	5.7	2597	10 US-09-905-129-13	Sequence 13, Appl
16	158.5	5.7	2597	10 US-09-991-630-2	Sequence 2, Appl1
17	158.5	5.7	2597	10 US-09-991-630-10	Sequence 10, Appl
18	158.5	5.7	2597	10 US-09-991-630-13	Sequence 13, Appl
19	157.5	5.7	941	12 US-10-124-557-14	Sequence 14, Appl

20	157.5	5.7	1022	12 US-10-124-557-84	Sequence 84, Appl
21	157.5	5.7	1038	12 US-10-124-557-74	Sequence 74, Appl
22	157.5	5.7	1049	12 US-10-124-557-58	Sequence 58, Appl
23	157.5	5.7	1140	12 US-10-124-557-104	Sequence 104, App
24	157.5	5.7	1270	12 US-10-124-557-44	Sequence 44, Appl
25	157.5	5.7	1311	12 US-10-124-557-42	Sequence 42, Appl
26	157.5	5.7	1313	12 US-10-124-557-142	Sequence 142, App
27	157.5	5.7	1314	12 US-10-124-557-50	Sequence 50, Appl
28	157.5	5.7	1320	12 US-10-124-557-46	Sequence 46, Appl
29	157.5	5.7	1354	12 US-10-124-557-48	Sequence 48, Appl
30	157.5	5.7	1361	12 US-10-124-557-40	Sequence 40, Appl
31	157.5	5.7	1363	12 US-10-124-557-52	Sequence 52, Appl
32	157.5	5.7	1404	12 US-10-124-557-2	Sequence 2, Appl1
33	157.5	5.7	1404	12 US-10-124-557-62	Sequence 62, Appl
34	157.5	5.7	1575	10 US-09-864-761-36047	Sequence 36047, A
35	157	5.7	2586	10 US-09-905-129-11	Sequence 11, Appl
36	156	5.7	2586	10 US-09-905-129-14	Sequence 14, Appl
37	156	5.7	2586	10 US-09-991-630-11	Sequence 11, Appl
38	156	5.7	2586	10 US-09-991-630-14	Sequence 14, Appl
39	156	5.7	2587	10 US-09-905-129-16	Sequence 16, Appl
40	156	5.7	2587	10 US-09-991-630-16	Sequence 16, Appl
41	156	5.7	2589	10 US-09-991-630-24	Sequence 24, Appl
42	156	5.7	2589	10 US-09-991-630-24	Sequence 24, Appl
43	155	5.6	2005	10 US-09-735-367B-3	Sequence 3, Appl1
44	155	5.6	2063	10 US-09-735-367B-2	Sequence 2, Appl1
45	153.5	5.6	2828	10 US-09-905-129-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-09-893-519A-14
Sequence 14, Application US/09893519A
Publication No. US20030027243A1
GENERAL INFORMATION: ANADYS PHARMACEUTICALS, INC.
APPLICANT: THOMPSON, Craig
APPLICANT: MOORE, Jeffrey
APPLICANT: BURDMAN, Ed T.
APPLICANT: BRADLEY, John
APPLICANT: DESILVA, Thamara
APPLICANT: HARRIS, Sandra
APPLICANT: KOMARNITSKY, Svetlana
APPLICANT: MOORE, Daniel
APPLICANT: MCCOY, Melissa
APPLICANT: SANDERSON, Karen
APPLICANT: HAO, Tariq
APPLICANT: ZHU, Shuhao
APPLICANT: LONG, Fan
TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
FILE REFERENCE: 0342/1G548-US2
CURRENT APPLICATION NUMBER: US/09/893, 519A
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/215,164
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: US 60/224,457
NUMBER OF SEQ ID NOS: 146
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 1023
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Corresponds to SEQ ID NO: 87
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Human Genbank/CAA72189
DATABASE ENTRY DATE: 1997-06-25
RELEVANT RESIDUES: (1)..(1023)

US-09-893-519A-14

Query Match 18.9%; Score 521; DB 9; Length 1023;
Best Local Similarity 31.7%; Pred. No. 6,5e-27;
Matches 168; Conservative 73; Mismatches 157; Indels 132; Gaps 22;

QY 13 PRVSSG-----PRLPAPQIVAVKAPRTTITQIPANILQIPGTVLIXNSGFLM 61
DB 414 PRATTSIGIRATLTPVLAAPLPP-----PQNPRTNIQ---NPQLPPGMVLVSENQQL 464
QY 62 LVSPQOATVR-----AETSNITSRPANPQVAKICIVPNSSSOLIKKAVTPKYL 115
DB 465 MI-PQALMOMQAHQAQPTMAPRATPTSPVQISIVQAPGFTIIR-QVTP-----518
QY 116 AQIGTTVVTTPKPSVQSAVPPTSVVTPGKPLNTVT--TLKPSLGASS-----TPS 168
DB 519 ----TTLIKOV---SOAQTVQPSATLQSRPGVQPOLVLCGAQOTASLGATAVAQCTPQ 571
QY 169 NE-PNLKAENSAVQINLSSTMLENYKKCKNLTAMLIKLCSSGSPSEKQNKVLEQL 227
DB 572 RIVPGATTTSSAATE-----TMENVKCKNLTSLIKLASSGKSTETANVKEVLQNL 625
QY 228 LPAKIAEETRLLYVELKSSQPHLVPELKSVALROLLPNSQSF10QCVQO---TS 283
DB 626 LQKIEADEPSTSLYRELNSSPQVILVPELKRSLPALROLTPDSANFIQSSQOQPPPTIS 665
QY 284 SDNVIACTTTVTSPVVTTVSSSQSEKSIIVSGATAPRTVSQVOTLNLPLAGPVAKAGV 343
DB 686 Q-----ATTALTAVLVSSSVQRTAGKTAATVTSALQPVLSL-----722
QY 344 VTLHSVGPRTAATGTAGCGLLOTSKRPLTVSVANTVTVTSLOPEKPVVSTAVTSLIPAV 403
DB 723 -----TQPTOVGKQOGOPPLVLIQ-----QPPKP-----GALIRPPOV 756
QY 404 TEGETSGAALCEPSVPSFCDHICKPVGTPVQIKLAQPGPVL---SOPAGIPPTS 459
DB 757 TLTGT-----PMVALRQPH-NRIMLTTPQOVNLSSEBARIATATSELVGLTIRS 804
QY 460 SSKQFLPSLEHVQ-----QPSGNEKQ---VTTTSHSSTLTITQKCGQK 499
DB 805 CKDEFLAPLQRLILEIGKKHGITELHPDVSVYSHATQORLQNLVER 854

RESULT 2

US-09-801-368-108
Sequence 108, Application US/09801368
Patent No. US20020128250A1
GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Cali, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Maxon, Mary
APPLICANT: Milne, Todd
APPLICANT: No. US20020128250A1man, Thea
APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amir
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801,368
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US 60/160,587
PRIOR FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 108
LENGTH: 1367

TYPE: PR1
ORGANISM: Saccharomyces cerevisiae
US-09-801-368-108

Query Match 8.0%; Score 220; DB 10; Length 1367;
Best Local Similarity 22.5%; Pred. No. 8.7e-07;
Matches 131; Conservative 77; Mismatches 240; Indels 134; Gaps 22;

QY 2 TLVTKAPVAPAPKPVSSGPRLPAP-----QIVAVKAPRTTITQIPANILQIPGTVLIX 54
DB 618 TTESSAPVTSSTTESSAPVPPTSSSTTESSAPVPTPSSSTTESSAPVPPTSSSTTE 677
QY 55 SNSGPEMLVSPQOT---VTAETSNITSRPANPQVAKICIVPNSSSOLIKKAV-- 109
DB 678 SSSAPVTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPV 737
QY 110 -TPVKKLAQIGTVVTV-----VPKPSVQS-----VAVPTSVVTVTGKPLNT- 152
DB 738 PTPSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTP 797
QY 153 -----VTLKPSLGASTPSENEPMLKAENSAVQINLSPTMLENVKCKNFLA 201
DB 798 SSSSTTESSAPVPTPSSSTNITSAPSTPSSSTTESSAPV---PVPSSSTTESSS--- 851
QY 202 MLIKLACSGSQS-----PEKQONKVLQQLDAKIEAETTRKLYVELKSSQPHLV 254
DB 852 --APVSSSTTESSAPVPTPSSSNTITSAPSSIPESSSTTESSSTGTVTPSSSK----- 904
QY 255 PFLKSVVALROLLPNSQSF10QCVQOQSSDNVIACTTIVTSP---VTTTVSSSQSE 311
DB 905 -----YPSGQT--ETSVSSTTETTVIPKTTISVTPSTTITTVVCSGTGN 949
QY 312 KSIIVSGATAPRTV--SVQTLNPLAGPVAKAGVTLHSVGPRTAATGTAGTG--TLQ 367
DB 950 SAGPTTSGCSPKVTVTPTTTTTSVTTSSTTITTVVCSGTGNAGETTSQSPKVTIT 1009
QY 368 SKPLVTSVANT---VTVSLOPEKPVVSGTAVTSLIPANT--FGETSGAALC----- 414
DB 1010 TVPSTSPSESTASSTTSPPTPVTVVSTTVVTEYSTKPGGEITTVFVKNIPPTV 1069
QY 415 -----LPSVKPVVPSFCDHICKPVGTPVQIKLAQPGVLSQPSGSSKQFLPSL 468
DB 1070 LTTIAPPTSVTITVNF-----TPTTITTT---YCS-----TGTNS----- 1101
QY 469 HVVQOPSGNEKQVTTTSHSSTLTITQKCGKTMVNTIPTIS 510
DB 1102 -AGETTSQSPKVTVTTVPCSTGT---GEVTEATTVLVTTA 1138

RESULT 3

US-10-025-380-1068
Sequence 1068, Application US/10025380
Publication No. US20020182191A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tongtong
APPLICANT: Jiang, Yugu
APPLICANT: Smith, Carole L.
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Skelky, Yasir A. W.
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick Thomas S.
APPLICANT: Carter, Derrick
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
FILE REFERENCE: 210121.471C14

```

Db 427 TEVTTTGTGNGQPPDEN-----VIVIRTPISGLITTTTE-----PWTGTFST 470
Qy 57 SGPLMLV-----SPQOT-----VTRAETTSNITSRPAPV-----ANPQVKTCTPVSSSOLI 104
Db 471 STEMTYTGNGQPTDEVIIVIRTPISGLITTTTEPWTGTFSTSTREVTITGTNGOPT 530
Qy 105 KKVAV-----TPYKLAQIGTIVTTVPKPSVQSAVPTSVVTPGKPLN-TVTTLKPS 160
Db 531 DEVIIVIRTPIS-----GLITTTTEPWTGTFSTSTREVTITGTNGOPTDEVIIVIRPT 586
Qy 161 LGASTTSNENPKAENSAVOI-----NISPTMLENKKCKNFAMLIK---LACSGS 211
Db 587 SEGILITTTTEPWTGTFSTSTREVTITGTNGOPT-----DEVIIVIRTPPTAISSS 637
Qy 212 QSPMGONVKKLVEQLDAKIEAEFTRKLYVELKSSPOPLVPE-----LKASV- 262
Db 638 LSSSGQITSSI-----TSSRPIITPFYBSNGSVISSVVIS 674
Qy 263 -ALRQLPNSQSF1QOCV-----OQTSMDVIACTTIVTTSPVTTVSSSQ 309
Db 675 SSYTSLSVTSSSFSSSVISSSTTSTISFSESSTSVIPTSSSGSEKTSASASS 734
Qy 310 SEKTIIVSGATAPRTVVOITNPLAGPVGAKAGVYTLHSVQPTAATGTTAGTGLQTSK 369
Db 735 SSSSISSSESPKSP--TNSSSSSLPV---TSATTGQETASSL--PPATTTKTSB---QTTL 784
Qy 370 PLATSVANTYTTVLSLOPEKPVSGTAVTLSLPATVFGTSGAALCLDSVKR----- 420
Db 785 VTVTSCSHCTESI--SSAIVSTAVTVS--GVTTETTTWCPISTETTTQOTGTTBOT 840
Qy 421 -----VSEFCMDHICKPVIGTPVOIKLAOPGVLSOPA---GIPV--- 457
Db 841 KGTTEQTTETTKQTTVTVTISCESDICS-----KTASPAIVSTATINQVTEYT 891
Qy 458 -----GSSSQQLFSLRHVVOQPSG--GNEKOVTTISHSSTLTITQCGCKTNPVTIIPTS 510
Db 892 TWCPDISTESKQOTLVTVTSCESGVCSETPSPAIVSTAVTAVN-----DVTYVPTW 944
Qy 511 QPPASIIKOTLLPGNKILSLQASPTOKNRKEVNTS 547
Db 945 R-PQTMQSVSSKMSNATSETTTNTGAAETKTAVTS 980

RESULT 11
US-09-801-368-106
; Sequence 106, Application US/09801368
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250Aiman, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 106

```

```

; LENGTH: 1169
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-106

Query Match      6.28; Score 172; DB 10; Length 1169;
Best Local Similarity 21.48; Pred. No. 0.0011;
Matches 112; Conservative 85; Mismatches 191; Indels 136; Gaps 24;

Qy 4 VTKAVPAPPKVSSGRLPAPQIVAVKAPMTTIQF--PAMQLPQGVILKNSGPTML 62
Db 689 VTSSSVSTPITSSESSASVTLI---PSTISEKPSWK---TKVVISSSPTNL 740
Qy 63 VSPQOTVTRAETTSNITSRPAPV-----PANPQVKTCTPVNSSSOLIKKAVAPV 112
Db 741 ITSYDTTNSKDTVSSSTSVSLISSISLPSYSKASBQDIFHSSIVSSGQALTFESSIKV 800
Qy 113 KK-----IAQIGTVYVTPK---PSSVQSAVPTSVVTPGPK 149
Db 801 SSSSESHRTSPYTSRSGIKSSGVELESTSTSFHFETSTASTVQISSQFPESSP 860
Qy 150 INTVTTTLKPSLSG---ASSPNEPNLKAENSAVOINLSPMLENKKCKNFAMLIK 205
Db 861 ISTVA---PSTGLNSQTESTNNSKETSSSENSASV----- 893
Qy 206 IACSSQSPENGQONVKKLVEQLDAKIEAE---FTR---KLYVELKSSPOPLVPELK 258
Db 894 MPSSSATPKTG-----KVISDETSSGFSRDTTYVRMTSETPTN---EQ 936
Qy 259 KSVVALROLPLNSQSF1QOCVQOQTSMDVIACTTVV-----SPVTTVSS 307
Db 937 TLLITVSSCESNCS-----NTVSSAVSTATTINGITTEVTTWCPLSATETLVSK 989
Qy 308 SOS-EKTIIVSGATAPRTVVOITNPL--LAPVGAAGVYTLHSVQPTAATGTTAGL 364
Db 990 LSEKTTTLITVTVTSCESGVCSETSPAIVSTAVTAVTVTSSVQATNKLAIVSSD- 1048
Qy 365 LQTSPLVTSVANTYTTVLSLOPEKPVSGTAVTLSLPATVFGTSGAALCLDSVKR 424
Db 1049 IENSKSKSPVSEAEETISIRNNFVP-TSGTISIEHT--TITSMSENSDNVS----- 1101
Qy 425 CWDHICKPVIGTPVOIKLAOPGVLSOPAGIPT---GSSSKOL 464
Db 1102 ASEAVSSKSVTNPVLSVSC-----OPRGTPASSMISSTASL 1139

RESULT 12
US-10-161-510-10
; Sequence 10, Application US/10161510
; Publication No. US20020192695A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: PIBS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-074C
; CURRENT APPLICATION NUMBER: US/10/161,510
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 60/296,076
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,605
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/357,253
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 1056
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-161-510-10

Query Match      5.88; Score 159.5; DB 9; Length 1056;
Best Local Similarity 21.58; Pred. No. 0.0064;
Matches 120; Conservative 56; Mismatches 210; Indels 171; Gaps 21;

```

Query Match	Similarity	5.7%	Score 158.5	DB 10	Length 2597
Best Local	Similarity 20.5%			Pred. No. 0.024	
Matches 143	Conservative 81		Mismatches 242	Indels 231	Gaps 31
QY	2	TLTVKAVAPVAD--PKVSGPRLPAPQIVAVKAPNTTIQF-----PANTOLPQGT-----	50		
Db	932	SPTSKASTVLOQINPTESGPIQIP--ITGVRPSSSDISHTTADPSPSSHSSHTTA	988		
QY	51	-----VLKNSGPRMLM---VSPQOTVTRAETTSN-----ITSRAV--PAMP	88		
Db	989	SSLFHIPRRNNTGNPPRLRHGERTIWSRGKVNKHRTPYLRRHRTVRAIIGPANK	1048		
QY	89	QTVKI-----CVPNSSOLIKV-----AVTPVK--	114		
Db	1049	NVSQVPAIEYEGMCHTCCSAEGYIATAALSPSSSHSALPRTNNGVIAESTTVAKP	1108		
QY	115	-----LQIGTVTVTVK-----PSVQSAVAVPTSV-----	142		
Db	1109	LLLFKQNVNDIEITTTTKKSGCESNHVIPLEASMTSAPTSVLCKSPVDNSGLSMPG	1168		
QY	143	TVTPEKPLNTVTTTLKPSSLGASSTPSNEPNLKAENSAVQINLSPTMLENVK-----	195		
Db	1169	TIQIGKD--SVETTPPSPSLSPSISPTSTKSKRKP-----LHQIFVNNOKKEGMLKNP	1221		
QY	196	-----CKPFLMLILYA---CGSQSPMGQVKKLVEQLLDKAEAEFTRKLYVELKS	247		
Db	1222	YQFGLQKNPAKLPKIALPLPTGSSPS-----DS-----TTLT	1256		
QY	248	SPQPHL-----VPFLKSVVALROLPLNSOSFIQOCVOOTSSDMVATCTT	293		
Db	1257	SPPPALSTPMATONKGTGEVVSAGARSLGAKKQPTTNSPVLPTSRMSTLNLSTET	1316		
QY	294	TVTTSPPVTTVSSSQSEKSLIVSGATAPRTVSVOTLNLPLAGPQAK--AGVTLHVSVP	351		
Db	1317	PTVSPPTASTVIMSEY-----RTRSKAKDOINGPRKRNRRNANTTPROVSGY	1365		
QY	352	TAATGGTAGAGTLQTSKPLTVSVANTVTVYSLOQEKPVSGTAY--TLSPAVY-----	404		
Db	1366	SAYSLMTIADPRLPASHSPRODDGN-----VSANVAHSTSLSLATELFEEK	1412		
QY	405	FGETSGAAILCPYKPVVSCWDHICKPVGISTPVOIKLADGPVLSQPA-GIPTGS----	459		
Db	1413	YTQIGLNTTALETTL-----LSKQSESTYVRASDTPPLLSGAPVPVPSPPEF	1463		
QY	460	-----SSQQLSLRPVVOQPSGGENKQYTTISHSS--TLTIQKCGQKMPVNTII-----	507		
Db	1464	TGCVVTDKVSASFOM-----TSNRVVTIYESRHTIDLOQPSAEASPEPEITGTDT	1516		
QY	508	PTSQFPASPI-LKOITLPGKKILSLQASPPQKNKIK	542		
Db	1517	SPSNLPFSTVPLARVXDPQNSKKMKPSBPWEHKIOLK	1553		
RESULT 14					
US-09-905-129-10					
Sequence 10, Application US/09905129					
Patent No. US20020137705A1					
GENERAL INFORMATION:					
APPLICANT: Elnat, et al					
TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THEREOF					
TITLE OF INVENTION: AND USFS THEREOF					
FILE REFERENCE: 540579-2007.2					
CURRENT APPLICATION NUMBER: US/09/905,129					
CURRENT FILING DATE: 2001-07-13					
PRIOR APPLICATION NUMBER: 09/802,318					
PRIOR FILING DATE: 2001-03-08					
PRIOR APPLICATION NUMBER: 60/207,821					
PRIOR FILING DATE: 2000-05-30					
PRIOR APPLICATION NUMBER: 60/084,944					
PRIOR FILING DATE: 1998-05-11					
PRIOR APPLICATION NUMBER: 60/085,673					
PRIOR FILING DATE: 1998-05-15					

;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 10
;; LENGTH: 2597
;; TYPE: PRT
;; ORGANISM: Rattus sp.
;; NAME/KEY: misc.feature
;; LOCATION: (1)..(2597)
;; OTHER INFORMATION: 'x' can be any amino acid
US-09-905-129-10

Query Match 5.7%; Score 158.5; DB 10; Length 2597;
Best Local Similarity 20.5%; Pred. No. 0.024;
Matches 143; Conservative 81; Mismatches 242; Indels 231; Gaps 31;

QY 2 TLVTAVAPVAP--PKVSSGRLPAQIVAVKAPVTITQF-----PANLQDPGT----- 50
D 932 SFTSKASTVLOPINTSEYGPQIP--ITGVSRSSDISHTTADPFSHSGSHHTA 988
QY 51 -----VLKNSGRLML--VSPQQTVAETTSN-----TTSRPV--PAMP 88
D 989 SLEFIIPNNMTGNFRLRHGRERTISRGVKNPHRTPYLRHRHRTVPRAIGPANK 1048
QY 89 QTVKI-----CTVNSSQLIKV-----ATPVKK- 114
D 1049 NVSQVATEYPMCHTCSABGLVATATAALSPSSSHSALPKTNVGVIAESTTVVKKP 1108
QY 115 -----LAQIGTVVTVPK-----PSSVQSVAVPTSVY----- 142
D 1109 LLKRDKNVDIEITTTTKTSGGESNHVITPEASMTAPTSVSGKSPVDNSGLSMGP 1168
QY 143 TVTPGKPLNTVTTLKPSISGASSTPSNEPNLKAENSAVQINLSPMLKNYK----- 195
D 1169 TIQTKD-SVETTPPLSPSLPTSPISPTSKFRKRP-----LHQIFVNNQKKEGLKNP 1221
QY 136 -----CKNFLAMLKLA---CSGSQSPMGONVKKLVEQLDARTEAEFTKRLVYELKS 247
D 1222 YQFGLQKNPAKLPKIALPLPTGQSSPS-----DS-----TTLLT 1256
QY 248 SPQPHL-----VPFLKKSVALROLPLNSQSFIOQCVQOTSSDMVATCTT 293
D 1257 SPPRLSTMTMATONKKGTEVVSAGASLSAGKQPTNSPVLPTISKRSNTLNLSLET 1316
QY 294 TVTSPVVTTVSSQSEKSIIVSGATAPRTVSVQTLNPLAGPVGAK--AGVYTLHVSVP 351
D 1317 PTVTSPTATASYIMSETO-----RTRSKAKQDQIGPKRRNNANNTPTQVSGY 1365
QY 352 TAAAGGTAGGGLQTSKPLTVSVANTVTVSLOPEKRVSGTAV--TLSLPAYT----- 404
D 1366 SAYSALTATADPLPLAFSHSPRODDGN-----VSAVAHSTSLATILELEK 1412
QY 405 FGETSGAICLPVSVKPVVVSFCMDHICKPVIGTQVQIKLQGPVLSQA--GIPTGS----- 459
D 1413 YTOILGNTTALETTL-----LSKQSESTYKARASDTPPLLSGAPVPPTSPPEF 1463
QY 460 -----SSKQLFSLFHVVOOPSGGNEKQVTTISHSS--TLTIQKCGQKMPVNTII----- 507
D 1464 TKGVVTDKVSAFQM-----TSNRVVTIYESSRHNTDLOQPSAEASPNPEITGTTD 1516
QY 508 -PTSOFPAST-LKQITLPKMKIISLQASPTQKNRIK 542
D 1517 SPSNLFPTSVPLKVDKPRONSKMKPSPWPERKHYOLK 1553

RESULT 15
US-09-905-129-13
; Sequence 13, Application US/09905129
; Patent No. US20020137705A1
; GENERAL INFORMATION:
; APPLICANT: Eliat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THERE
; FILE REFERENCE: 540579-2007.2

;; CURRENT APPLICATION NUMBER: US/09/905,129
;; CURRENT FILING DATE: 2001-07-13
;; PRIOR APPLICATION NUMBER: 09/802,318
;; PRIOR FILING DATE: 2001-03-08
;; PRIOR APPLICATION NUMBER: 60/207,821
;; PRIOR FILING DATE: 2000-05-30
;; PRIOR APPLICATION NUMBER: 60/084,944
;; PRIOR FILING DATE: 1998-05-11
;; PRIOR APPLICATION NUMBER: 60/085,673
;; PRIOR FILING DATE: 1998-05-15
;; NUMBER OF SEQ ID NOS: 25
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 13
;; LENGTH: 2597
;; TYPE: PRT
;; ORGANISM: Rattus sp.
;; NAME/KEY: misc.feature
;; LOCATION: (1)..(2597)
;; OTHER INFORMATION: 'x' can be any amino acid
US-09-905-129-13

Query Match 5.7%; Score 158.5; DB 10; Length 2597;
Best Local Similarity 20.5%; Pred. No. 0.024;
Matches 143; Conservative 81; Mismatches 242; Indels 231; Gaps 31;

QY 2 TLVTAVAPVAP--PKVSSGRLPAQIVAVKAPVTITQF-----PANLQDPGT----- 50
D 932 SFTSKASTVLOPINTSEYGPQIP--ITGVSRSSDISHTTADPFSHSGSHHTA 988
QY 51 -----VLKNSGRLML--VSPQQTVAETTSN-----TTSRPV--PAMP 88
D 989 SLEFIIPNNMTGNFRLRHGRERTISRGVKNPHRTPYLRHRHRTVPRAIGPANK 1048
QY 89 QTVKI-----CTVNSSQLIKV-----ATPVKK- 114
D 1049 NVSQVATEYPMCHTCSABGLVATATAALSPSSSHSALPKTNVGVIAESTTVVKKP 1108
QY 115 -----LAQIGTVVTVPK-----PSSVQSVAVPTSVY----- 142
D 1109 LLKRDKNVDIEITTTTKTSGGESNHVITPEASMTAPTSVSGKSPVDNSGLSMGP 1168
QY 143 TVTPGKPLNTVTTLKPSISGASSTPSNEPNLKAENSAVQINLSPMLKNYK----- 195
D 1169 TIQTKD-SVETTPPLSPSLPTSPISPTSKFRKRP-----LHQIFVNNQKKEGLKNP 1221
QY 136 -----CKNFLAMLKLA---CSGSQSPMGONVKKLVEQLDARTEAEFTKRLVYELKS 247
D 1222 YQFGLQKNPAKLPKIALPLPTGQSSPS-----DS-----TTLLT 1256
QY 248 SPQPHL-----VPFLKKSVALROLPLNSQSFIOQCVQOTSSDMVATCTT 293
D 1257 SPPRLSTMTMATONKKGTEVVSAGASLSAGKQPTNSPVLPTISKRSNTLNLSLET 1316
QY 294 TVTSPVVTTVSSQSEKSIIVSGATAPRTVSVQTLNPLAGPVGAK--AGVYTLHVSVP 351
D 1317 PTVTSPTATASYIMSETO-----RTRSKAKQDQIGPKRRNNANNTPTQVSGY 1365
QY 352 TAAAGGTAGGGLQTSKPLTVSVANTVTVSLOPEKRVSGTAV--TLSLPAYT----- 404
D 1366 SAYSALTATADPLPLAFSHSPRODDGN-----VSAVAHSTSLATILELEK 1412
QY 405 FGETSGAICLPVSVKPVVVSFCMDHICKPVIGTQVQIKLQGPVLSQA--GIPTGS----- 459
D 1413 YTOILGNTTALETTL-----LSKQSESTYKARASDTPPLLSGAPVPPTSPPEF 1463
QY 460 -----SSKQLFSLFHVVOOPSGGNEKQVTTISHSS--TLTIQKCGQKMPVNTII----- 507
D 1464 TKGVVTDKVSAFQM-----TSNRVVTIYESSRHNTDLOQPSAEASPNPEITGTTD 1516
QY 508 -PTSOFPAST-LKQITLPKMKIISLQASPTQKNRIK 542
D 1517 SPSNLFPTSVPLKVDKPRONSKMKPSPWPERKHYOLK 1553

Thu Feb 20 16:36:32 2003

us-09-763-909-2_copy_1_552.rapb

Page 10

Search completed: February 16, 2003, 22:02:06
Job time : 39.8745 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 22:01:05 ; Search time 194.796 Seconds
(without alignments)
1827.005 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_552

Perfect score: 2758
Sequence: 1 GTLVTKVAPVAPKVSQGP.....ASPTOKRIKENTSCFRDE 552

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending_Patents_AA_Main:*

1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US097_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US099_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US102_COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US103_COMB.pep.*
27: /cgn2_6/ptodata/2/paa/US104_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2758	100.0	801	1	PCT-US01-08631-39827
2	2758	100.0	852	21	US-09-763-909-2
3	2678	97.1	865	1	PCT-US02-25829-25
4	2665	96.6	843	1	PCT-US02-29964-410
5	2235	81.0	685	27	US-60-243-468-1271
6	547	19.8	737	1	PCT-US94-01114-16

7	547	19.8	1051	1	PCT-US02-30474-1660	Sequence 1660, Ap
8	547	19.8 <th>1051</th> <th>27</th> <th>US-60-324-631-1665</th> <th>Sequence 1665, Ap</th>	1051	27	US-60-324-631-1665	Sequence 1665, Ap
9	535	19.4 <th>1083</th> <th>1</th> <th>PCT-US02-30474-3271</th> <th>Sequence 3271, Ap</th>	1083	1	PCT-US02-30474-3271	Sequence 3271, Ap
10	521	18.9 <th>1023</th> <th>1</th> <th>PCT-US01-20592-14</th> <th>Sequence 14, Ap</th>	1023	1	PCT-US01-20592-14	Sequence 14, Ap
11	521	18.9 <th>1023</th> <th>1</th> <th>PCT-US01-20592A-14</th> <th>Sequence 14, Ap</th>	1023	1	PCT-US01-20592A-14	Sequence 14, Ap
12	521	18.9 <th>1023</th> <th>22</th> <th>US-09-893-519A-14</th> <th>Sequence 14, Ap</th>	1023	22	US-09-893-519A-14	Sequence 14, Ap
13	258	9.4 <th>899</th>	899	27	US-60-167-217-11401	Sequence 11401, A
14	258	9.4 <th>921</th>	921	27	US-60-173-464-9223	Sequence 9223, Ap
15	258	9.4 <th>921</th>	921	27	US-60-173-464-9223	Sequence 9223, Ap
16	256	9.3 <th>921</th>	921	1	PCT-US94-01114-2	Sequence 20611, A
17	256	9.3 <th>921</th>	921	4	US-08-013-412-2	Sequence 2, Ap
18	256	9.3 <th>921</th>	921	20	US-09-614-150-11376	Sequence 11376, A
19	256	9.3 <th>921</th>	921	20	US-09-614-150-24957	Sequence 24957, A
20	256	9.3 <th>921</th>	921	27	US-60-191-637-11408	Sequence 11408, A
21	256	9.3 <th>921</th>	921	27	US-60-191-637-25058	Sequence 25058, A
22	256	9.3 <th>921</th>	921	27	US-60-191-681-8933	Sequence 8933, Ap
23	256	9.3 <th>921</th>	921	27	US-60-191-681-19739	Sequence 19739, A
24	247.5	9.0 <th>1953</th> <th>27</th> <th>US-60-185-361-566</th> <th>Sequence 566, App</th>	1953	27	US-60-185-361-566	Sequence 566, App
25	247.5	9.0 <th>2947</th> <th>27</th> <th>US-60-185-361-494</th> <th>Sequence 494, App</th>	2947	27	US-60-185-361-494	Sequence 494, App
26	247.5	9.0 <th>2947</th> <th>27</th> <th>US-60-185-361-938</th> <th>Sequence 938, App</th>	2947	27	US-60-185-361-938	Sequence 938, App
27	237	8.6 <th>1149</th> <th>27</th> <th>US-60-185-361-565</th> <th>Sequence 565, App</th>	1149	27	US-60-185-361-565	Sequence 565, App
28	233.5	8.5 <th>528</th> <th>22</th> <th>US-09-840-746-20</th> <th>Sequence 20, App</th>	528	22	US-09-840-746-20	Sequence 20, App
29	224	8.1 <th>564</th> <th>18</th> <th>US-09-417-507-42234</th> <th>Sequence 42234, A</th>	564	18	US-09-417-507-42234	Sequence 42234, A
30	220.5	8.0 <th>4742</th> <th>26</th> <th>US-10-221-279-12349</th> <th>Sequence 12349, A</th>	4742	26	US-10-221-279-12349	Sequence 12349, A
31	220	8.0 <th>1367</th> <th>18</th> <th>US-09-487-558-108</th> <th>Sequence 108, App</th>	1367	18	US-09-487-558-108	Sequence 108, App
32	220	8.0 <th>1367</th> <th>22</th> <th>US-09-487-558-108</th> <th>Sequence 108, App</th>	1367	22	US-09-487-558-108	Sequence 108, App
33	220	8.0 <th>1367</th> <th>22</th> <th>US-09-801-368-108</th> <th>Sequence 108, App</th>	1367	22	US-09-801-368-108	Sequence 108, App
34	220	8.0 <th>1367</th> <th>27</th> <th>US-60-087-236-55</th> <th>Sequence 55, App</th>	1367	27	US-60-087-236-55	Sequence 55, App
35	219	7.9 <th>695</th> <th>27</th> <th>US-09-641-377-694</th> <th>Sequence 694, App</th>	695	27	US-09-641-377-694	Sequence 694, App
36	219	7.9 <th>745</th> <th>20</th> <th>US-09-641-377-694</th> <th>Sequence 694, App</th>	745	20	US-09-641-377-694	Sequence 694, App
37	219	7.9 <th>745</th> <th>20</th> <th>US-09-641-377-696</th> <th>Sequence 696, App</th>	745	20	US-09-641-377-696	Sequence 696, App
38	219	7.9 <th>5179</th>	5179	19	PCT-US02-17382-218	Sequence 218, App
39	219	7.9 <th>5179</th>	5179	19	US-09-538-092-1258	Sequence 1258, Ap
40	219	7.9 <th>5179</th>	5179	20	US-09-649-811-1068	Sequence 1068, Ap
41	219	7.9 <th>5179</th>	5179	20	US-09-649-811-1068	Sequence 1068, Ap
42	219	7.9 <th>5179</th>	5179	22	US-09-833-263-1068	Sequence 1068, Ap
43	219	7.9 <th>5179</th>	5179	23	US-09-922-217-1068	Sequence 1068, Ap
44	219	7.9 <th>5179</th>	5179	24	US-10-025-380-1068	Sequence 1068, Ap
45	216	7.8 <th>778</th> <th>20</th> <th>US-09-641-377-685</th> <th>Sequence 685, App</th>	778	20	US-09-641-377-685	Sequence 685, App

ALIGNMENTS

RESULT 1
PCT-US01-08631-39827
Sequence 39827, Application PC/TUS0108631
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc
TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
FILE REFERENCE: 21272-049
CURRENT APPLICATION NUMBER: PCT/US01/08631
CURRENT FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: 09/540,217
PRIOR FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: 09/649,167
PRIOR FILING DATE: 2000-08-23
NUMBER OF SEQ ID NOS: 60736
SOFTWARE: Custom
SEQ ID NO 39827
LENGTH: 801
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US01-08631-39827

Query Match 100.0%; Score 2758; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 2, 2e-215; Indels 0; Gaps 0;
Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTLVTKVAPVAPKVSQGPLPAPQIVAVKAPMTTITOPANQLDPGTVLKSNSGPL 60
DB 1 GTLVTKVAPVAPKVSQGPLPAPQIVAVKAPMTTITOPANQLDPGTVLKSNSGPL 60

;; TITLE OF INVENTION: NUCLEIC ACID ASSOCIATED PROTEINS
;; FILE REFERENCE: PF-1146 PCT
;; CURRENT APPLICATION NUMBER: PCT/US02/25829
;; PRIOR FILING DATE: 2002-08-14
;; PRIOR APPLICATION NUMBER: US 60/313,111
;; PRIOR FILING DATE: 2001-08-17
;; PRIOR APPLICATION NUMBER: US 60/314,682
;; PRIOR FILING DATE: 2001-08-24
;; PRIOR APPLICATION NUMBER: US 60/314,756
;; PRIOR FILING DATE: 2001-08-24
;; PRIOR APPLICATION NUMBER: US 60/315,105
;; PRIOR FILING DATE: 2001-08-27
;; PRIOR APPLICATION NUMBER: US 60/316,751
;; PRIOR FILING DATE: 2001-08-31
;; PRIOR APPLICATION NUMBER: US 60/316,856
;; PRIOR FILING DATE: 2001-08-31
;; PRIOR APPLICATION NUMBER: US 60/328,185
;; PRIOR FILING DATE: 2001-10-05
;; NUMBER OF SEQ ID NOS: 66
;; SOFTWARE: PERL Program
;; SEQ ID NO 25
;; LENGTH: 865
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc.feature
;; OTHER INFORMATION: Incyte ID No: 4398735CD1
PCT-US02-25829-25

Query Match 97.1%; Score 2678; DB 1; Length 865;
Best Local Similarity 97.7%; Pred. No. 8.3e-209;
Matches 542; Conservative 0; Mismatches 5; Indels 8; Gaps 2;

OY 1 GTLVTKVAPVAPPVSSGPRLPAPQIVAVKAPMTTIIQFPANLQLP---GTVLIRISNS 57
DB 67 GTLVTKVAPVAPPVSSGPRLPAPQIVAVKAPMTTIIQFPANLQLP---GTVLIRISNS 126
OY 58 GFLMLVSPQOQVTRAEFTSNITSRPAVPANQVYKICTVPSNSQLIKKVAVTPVKKLAQ 117
DB 127 GFLMLVSPQOQVTRAEFTSNITSRPAVPANQVYKICTVPSNSQLIKKVAVTPVKKLAQ 186
OY 118 IGTVVTVTRKPSVQAVAPTSVYVTPGKPLNTVTTLKPSISGASSTPNEPNLKAEN 177
DB 187 IGTVVTVTRKPSVQAVAPTSVYVTPGKPLNTVTTLKPSISGASSTPNEPNLKAEN 246
OY 178 SAAVOIMLSPMLENVKCKNFMLMLIKIACSGSOSPEMGONVKKIVEQLDAKIEAEF 237
DB 247 SAAVOIMLSPMLENVKCKNFMLMLIKIACSGSOSPEMGONVKKIVEQLDAKIEAEF 306
OY 238 TRKLYELKSSPOHLVPLFKKSVVALROLPLNSQSFIOQCVQOTSSDMVATCTTYYT 297
DB 307 TRKLYELKSSPOHLVPLFKKSVVALROLPLNSQSFIOQCVQOTSSDMVATCTTYYT 366
OY 298 SPVYTTVVSSQSKSIIVSAGATAPRIVSVQTLPLAGPVGAKGVYTLHSVGPAAATGG 357
DB 367 SPVYTTVVSSQSKSIIVSAGATAPRIVSVQTLPLAGPVGAKGVYTLHSVGPAAATGG 426
OY 358 TTAGTGLQTSKPLVTSVANTVTVSLOPEKPVVSGTAVTLTLAVYFGEESGAICLPS 417
DB 427 TTAGTGLQTSKPLVTSVANTVTVSLOPEKPVVSGTAVTLTLAVYFGEESGAICLPS 486
OY 418 VKPVVSCWMDHICKPVIGTPOIKLAQPGPYLSQPAIGTPTSSSKOLFSLFHVWQPSGG 477
DB 487 VKPVVSCWMDHICKPVIGTPOIKLAQPGPYLSQPAIGTPTSSSKOLFSLFHVWQPSGG 541
OY 478 NEKVYTISSSTLTIOCKGCKMPVNTIIPTSQPPASIIKQITLPGNKILSLQASPTQ 537
DB 542 NEKVYTISSSTLTIOCKGCKMPVNTIIPTSQPPASIIKQITLPGNKILSLQASPTQ 601
OY 538 KNRIKENVTSQCFRDE 552
DB 602 KNRIKENVTSQCFRDE 616

RESULT 4
PCT-US02-29964-410
Sequence 410, Application PC/TUS0229964
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Ren, Felyan
APPLICANT: Zhang, Jie
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Jian-Rui
APPLICANT: Xue, Aidong J.
APPLICANT: Wang, Dunrui
APPLICANT: Ghosh, Malabika
APPLICANT: Asundi, Vinod
APPLICANT: Wehrman, Tom
APPLICANT: Zhou, Ping
APPLICANT: Goodrich, Ryle W.
APPLICANT: Weng, Gezhi
APPLICANT: Haley-Vicente, Dana
APPLICANT: Drmanac, Radoje T
TITLE OF INVENTION: Novel Nucleic Acids and
FILE REFERENCE: 809ACIP PCT
CURRENT APPLICATION NUMBER: PCT/US02/29964
CURRENT FILING DATE: 2002-09-19
PRIOR APPLICATION NUMBER: US 60/323,739
PRIOR FILING DATE: 2001-09-19
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02623
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: US 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: PCT/US01/03800
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 09/515,126
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 922
SOFTWARE: PL_FL_genes Version 6.0
SEQ ID NO 410
LENGTH: 843
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US02-29964-410
Query Match 96.6%; Score 2665; DB 1; Length 843;
Best Local Similarity 97.8%; Pred. No. 9.2e-208;
Matches 540; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 1 GTLVTKVAPVAPPVSSGPRLPAPQIVAVKAPMTTIIQFPANLQLP---GTVLIRISNSGPT 60
DB 43 GTLVTKVAPVAPPVSSGPRLPAPQIVAVKAPMTTIIQFPANLQLP---GTVLIRISNSGPT 102
OY 61 MLVSPQOQVTRAEFTSNITSRPAVPANQVYKICTVPSNSQLIKKVAVTPVKKLAQIGT 120
DB 103 MLVSPQOQVTRAEFTSNITSRPAVPANQVYKICTVPSNSQLIKKVAVTPVKKLAQIGT 162
OY 121 TVVTVTRKPSVQAVAPTSVYVTPGKPLNTVTTLKPSISGASSTPNEPNLKAENSA 180
DB 163 TVVTVTRKPSVQAVAPTSVYVTPGKPLNTVTTLKPSISGASSTPNEPNLKAENSA 222
OY 181 VOIMLSPMLENVKCKNFMLMLIKIACSGSOSPEMGONVKKIVEQLDAKIEAEFTRK 240
DB 223 VOIMLSPMLENVKCKNFMLMLIKIACSGSOSPEMGONVKKIVEQLDAKIEAEFTRK 282

QY	241	LYVAKSSPOHLPVLFKKSVALROLPLPNSQSTQOCVOQTSSDMVIAICTTIVTTSV	300
Db	283	LYVELKSSPOHLPVLFKKSVALROLPLPNSQSTQOCVOQTSSDMVIAICTTIVTTSV	342
QY	301	VTTTVSSSQSEKSIIVSGATAPRTVSQOTLPLAGPVGAKAGVTTLSVGPATAIGTTA	360
Db	343	VTTTVSSSQSEKSIIVSGATAPRTVSQOTLPLAGPVGAKAGVTTLSVGPATAIGTTA	402
QY	361	GTGLQTSKPLVTSVANTVTTVLSLOPEKPVVSGTAVTLSLPAVTFGTSGAICLP	420
Db	403	GTGLQTSKPLVTSVANTVTTVLSLOPEKPVVSGTAVTLSLPAVTFGTSGAICLP	462
QY	421	VVSFCWDHICKPVYGTVPQIKLAOPGPVLSOPACIPYSSSKOLSLFHHVQDPSGNEK	480
Db	463	VVSAGTSDPKPVYGTVPQIKLAOPGPVLSOPACIPQAVQVKOLFSLFHHVQDPSGNEK	522
QY	481	QVTTTSHSSTLTICKCGKQKTMVNTIITPSQFPFASILIKQITLPGNKILLSQASPTQKNR	540
Db	523	QVTTTSHSSTLTICKCGKQKTMVNTIITPSQFPFASILIKQITLPGNKILLSQASPTQKNR	582
QY	541	IKENYVTSQFRDE	552
Db	583	IKENYVTSQFRDE	594

```

RESULT 5
US-60-243-468-1271
: Sequence 1271, Application US/60243468
: GENERAL INFORMATION:
: APPLICANT: Beasley, Ellen
: TITLE OF INVENTION: ISOLATED HUMAN DRUG TARGET PROTEINS,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN DRUG TARGET PROTEINS.
: TITLE OF INVENTION: AND USES THEREOF
: FILE REFERENCE: CL000929
: CURRENT APPLICATION NUMBER: US/60/243,468
: CURRENT FILING DATE: 2000-10-27
: NUMBER OF SEQ ID NOS: 2121
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1271
: LENGTH: 685
: TYPE: PRT
: ORGANISM: HUMAN
US-60-243-468-1271

```

Query Match	81.0%;	Score 2235;	DB 27;	Length 685;
Best Local Similarity	86.4%;	Pred. No. 8.5e-173;		
Matches 465;	Conservative 6;	Mismatches 23;	Indels 42;	Gaps 3
QY	1	GLVLTWKVAPVAPSPKVVSSGCPRLPAPQIYAVKAPNNTTIIQFPANLQLPPTVLIKNSGPL	60	
Db	67	GLVLTWKVAPVAPSPKVVSSGCPRLPAPQIYAVKAPNNTTIIQFPANLQLPPTVLIKNSGPL	126	
QY	61	MLVSPQOQVYTRAEETTSNTTSSRPVAPNPOYVAKICIVPNSSQLIKKVAVTPVKKLAQIGT	120	
Db	127	MLVSPQOQVYTRAEETTSNTTSSRPVAPNPOYVAKICIVP-----	163	
QY	121	TVVTVTPKPVSSQVAVPSTVYTPGKRLNVTTLKPSLIGASSTPSNPNLKAENSA	180	
Db	164	-----SVAVPSTVYTPGKRLNVTTLKPSLIGASSTPSNPNLKAENSA	210	
QY	181	VOINLSPMLLENVKKCKNFLAMLKILACSGSGSPENGGVKKLVEQLDAKTEAEFTTK	240	
Db	211	VOINLSPMLLENVKKCKNFLAMLKILACSGSGSPENGGVKKLVEQLDAKTEAEFTTK	270	
QY	241	LYVELKSPQPLVLPFLKKSVVALNQLLPNSOSFTIOQCVOQCTSSDMVIACTCTTVTTSV	300	
Db	271	LYVELKSPQPLVLPFLKKSVVALNQLLPNSOSFTIOQCVOQCTSSDMVIACTCTTVTTSV	330	
QY	301	VTTTVSSSSEKSIIVSGATAPRTVSYQTLNPLAGVGAAGVWVLHSGPRAAGGTTA	360	
Db	331	VTTTVSSSSEKSIIVSGATAPRTVSYQTLNPLAGVGAAGVWVLHSGPRAAGGTTA	390	

[illegible]

```

RESULT 6
PCT-US94-01114-16
: Sequence 16, Application PC/TUS9401114
: GENERAL INFORMATION:
: APPLICANT: Tjian, Robert
: APPLICANT: Comai, Lucio
: APPLICANT: Dynlacht, Brian D.
: APPLICANT: Hoey, Timothy
: APPLICANT: Ruppert, Siegfried
: APPLICANT: Tanese, Noko
: APPLICANT: Wang, Edith
: APPLICANT: Weinzierl, Robert O.J.
: TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
: TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TARS AND METHODS OF USE
: NUMBER OF SEQUENCES: 36
: CORRESPONDENCE ADDRESS:

```

ADDRESS: EMMETT HONANCK, INST. REGISTRATION & RESEARCH
STREET: 4 UNDERWOOD Centro, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCP/US94/01114
FILING DATE: 28-JAN-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: FP-57650-2/AUT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 737 amino acids
type: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01114-16

```

Query Match          19.8%; Score 547; DB 1; Length 737;
Best Local Similarity 30.2%; Pred. No. 6.4e-35;
Matches 173; Conservative 67; Mismatches 144; Indels 188; Gaps 21.

QY 13 PPKVSSG-----PRLPAPDIVAKAPNTTIOFPNALQPPGVLIKNSGPI 61
      | :|| | | | | | | | | | | | | | | | | | | | | |
Db 68 PTTATSSGIRATLTPYVLAERLPQ-----PQNPNINQ---NFQLPPEGVAVLRSENGQL 118

QY 62 LVSPQOIVTR-----AETTSNITSRPANVPANPQTVKICTVPNSSOLIKRVAVPVKLT 115
      : : | | : : : : | | | | | | | | | | | | | | | |
Db 119 MI-PQQAALQAQQAQAHAPQPTTMABRPANPTTSAPRPVQISTVGAQPTPIIAR-QVTP---- 172

QY 116 AQLGTTVVTTVPKRVSSVQSVAVAPVTSVTVTPDKPLNTVT--TLKDDSLGASS-----TPS 168

```

```

Db 173 -----TTIKOV---SQAQTVPQATLQRSFGVQPOLVLGGAQTALGATAVGTGPQ 225
QY 169 NE-PNLKAENSAVOINLSPTMLENKKCKNFMLIKLACSGSQSPMGONVKKLVEOL 227
Db 226 RTVGATTSSAATE-----TMENVKCKNFSTLTKLASSGQSTETANVKELONL 279
QY 228 LDKAIEAEFTKRLVLELKSSPOPHVLPFLKKSVALROLPLNSQSFIOOCVOO-----TS 283
Db 280 LDKGIEAEDEFTSRLYRELNSPPQYLVPLFKRSLPALROLTPDPAALIQSQOQPPPTTS 339
QY 284 SDVIAATCTTIVTSPVTTTSSQSEKSIYSGATAPRTVSQTLNPLAGPVGAKAGV 343
Db 340 Q-----ATTALTAVALSSSVORTAGTATATVSALQPVLSL----- 376
QY 344 VTLHSGVPTAATGCTAGTGLQTSKPLVYTSVANTVTVTSLOPEKPVVSGTAVTSLP 403
Db 377 -----TQPTQVGVKGQGPPLVIO-----QPKP-----GALIRPPQV 410
QY 404 TFGTSGAATCLPSVKPVVSCMDHICKPVYIGTPVQIKLAOPGVLSQAPAGIPGSSSKQ 463
Db 411 TLRTQ-----PMVALROPH-NRIMLTTPQOIQ----- 437
QY 464 LFSLFHVQOPSGGNEKQVTTISHSSTLTIOKCGQKTMPTVTTIPTSOPFPASTLKQITL 523
Db 438 -----NPLQPVVVKPAVL 451
QY 524 PGNKIL---SLQASPTQKNRIKENVTSCFRDE 552
Db 452 PGTALSAVSNQAAAAQKNKKEPGGGSFRD 483

```

RESULT 7

```

PCT-US02-30474-1660
; Sequence 1660, Application PC/TUS0230474
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhao, Qiang A.
; APPLICANT: Wang, Jien-Rui
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Ghosh, Malabika
; APPLICANT: Wang, Dunrul
; APPLICANT: Ma, Yundng
; APPLICANT: Asundi, Vinod
; APPLICANT: Wang, Zhiwei
; APPLICANT: Haley-Vicente, Dana
; APPLICANT: Drmanac, Radoje T
; TITLE OF INVENTION: Novel Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: PCT/US02/30474
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 60/324,631
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27

```

```

; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 3476
; SOFTWARE: PL-FL-Genes Version 6.0
; SEQ ID NO 1660
; LENGTH: 1051
; TYPE: PRT
; ORGANISM: Homo sapiens
; PCT-US02-30474-1660

```

```

Query Match 19.8%; Score 547; DB 1; Length 1051;
Best Local Similarity 30.2%; Pred. No. 1,1e-34;
Matches 173; Conservative 67; Mismatches 144; Indels 188; Gaps 21;

```

```

QY 13 PPKVSG-----PLPAPQIVAKAPNTTTOFPANLQPPGTVLIRKNSGGLM 61
Db 382 PTATSGIRATILPTVLAPRLPQ-----PQNPNTIQ-----NFQLPGMVLVRENGQL 432
QY 62 LVSPOQIVTR-----AETSNITSRPAVAPNPQTVACIYIPNSSSOLIKKVAATPVKKL 115
Db 433 MI-POQALQAOQAOAHQAPQTTMAPRATPTSAAPVQISTVQAPGTPIAR-QVTP----- 486
QY 116 AQIGTVTVTPPKPSVQAVAPTSVTVTPGKPLNTVT--TLKPSISGASS-----TPS 168
Db 487 -----TTIKOV---SQAQTVPQATLQRSFGVQPOLVLGGAQTALGATAVGTGPQ 539
QY 169 NE-PNLKAENSAVOINLSPTMLENKKCKNFMLIKLACSGSQSPMGONVKKLVEOL 227
Db 540 RTVGATTSSAATE-----TMENVKCKNFSTLTKLASSGQSTETANVKELONL 593
QY 228 LDKAIEAEFTKRLVLELKSSPOPHVLPFLKKSVALROLPLNSQSFIOOCVOO-----TS 283
Db 594 LDKGIEAEDEFTSRLYRELNSPPQYLVPLFKRSLPALROLTPDPAALIQSQOQPPPTTS 653
QY 284 SDVIAATCTTIVTSPVTTTSSQSEKSIYSGATAPRTVSQTLNPLAGPVGAKAGV 343
Db 654 Q-----ATTALTAVALSSSVORTAGTATATVSALQPVLSL----- 690
QY 344 VTLHSGVPTAATGCTAGTGLQTSKPLVYTSVANTVTVTSLOPEKPVVSGTAVTSLP 403
Db 691 -----TQPTQVGVKGQGPPLVIO-----QPKP-----GALIRPPQV 724
QY 404 TFGTSGAATCLPSVKPVVSCMDHICKPVYIGTPVQIKLAOPGVLSQAPAGIPGSSSKQ 463
Db 725 TLRTQ-----PMVALROPH-NRIMLTTPQOIQ----- 751
QY 464 LFSLFHVQOPSGGNEKQVTTISHSSTLTIOKCGQKTMPTVTTIPTSOPFPASTLKQITL 523
Db 752 -----NPLQPVVVKPAVL 765
QY 524 PGNKIL---SLQASPTQKNRIKENVTSCFRDE 552
Db 766 PGTALSAVSNQAAAAQKNKKEPGGGSFRD 797

```

RESULT 8

```

; Sequence 1665, Application US/60324631
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhao, Qiang A.
; APPLICANT: Wang, Jien-Rui
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Ghosh, Malabika
; APPLICANT: Wang, Dunrul
; APPLICANT: Ma, Yundng

```

APPLICANT: Asundi, Vinod
APPLICANT: Wang, Zhiwei
APPLICANT: Wang, Gezhi
APPLICANT: Haley-Vicente, Dana
APPLICANT: Drmanac, Radoje T
TITLE OF INVENTION: Novel Nucleic Acids and
FILE REFERENCE: Polypeptides
FILE REFERENCE: 810
CURRENT APPLICATION NUMBER: US/60/324,631
CURRENT FILING DATE: 2001-09-24
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02623
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: US 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: PCT/US01/03800
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 09/515,126
PRIOR FILING DATE: 2000-02-28
PRIOR APPLICATION NUMBER: US 09/577,409
PRIOR FILING DATE: 2000-05-18
PRIOR APPLICATION NUMBER: PCT/US01/04927
PRIOR FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: US 09/519,705
PRIOR FILING DATE: 2000-03-07
PRIOR APPLICATION NUMBER: US 09/574,454
PRIOR FILING DATE: 2000-05-19
PRIOR APPLICATION NUMBER: PCT/US01/04941
PRIOR FILING DATE: 2001-03-05
PRIOR APPLICATION NUMBER: US 09/540,217
PRIOR FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: US 09/649,167
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: PCT/US01/08631
PRIOR FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: US 09/552,929
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: US 09/770,160
PRIOR FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: PCT/US01/08656
PRIOR FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: US 09/577,408
PRIOR FILING DATE: 2000-05-18
PRIOR APPLICATION NUMBER: PCT/US01/14827
PRIOR FILING DATE: 2001-05-16
NUMBER OF SEQ ID NOS: 3334
SOFTWARE: pf_fl_genes Version 6.0
SEQ ID NO 1665
LENGTH: 1051
TYPE: PRT
ORGANISM: Homo sapiens
US-60-324-631-1665
Query Match 19.8%; Score 547; DB 27; Length 1051;
Best Local Similarity 30.2%; Pred. No. 1,1e-34;
Matches 173; Conservative 67; Mismatches 144; Indels 188; Gaps 21;
QY 13 PKVSSG-----RLPAPQVANKAPNTTIOFPAHQIAPPGVILKSNGLM 61
DB 382 PATTSGIRATLTPTVLABRLPOP-----PONPTIQ---NFQLEPGVAVLSENGQL 432
QY 62 LVSPQQTIVR-----AETTSNITSRPAPVAPNQTIVKICTVNSSQLKKVAVTPVKLT 115
DB 433 MI-PQQAALMQQAHAQQTMTMAPRPATPTSAPVQISTVQAGPPIIAR-QVTP----- 486

QY 116 AQIGTVVTVTPKPSVQSVANFTSVVTVPGKPLTV--TLKPSGLASS-----TPS 168
DB 487 ---TTIKQV---SQAQTTVQPSATLQSPGQPOLVYGAAQATSLGTATAVATGTGPQ 539
QY 169 NE-PNKAENSAVAOINLPTMLENKKCNKFNAMILIKLAGSGSOSPEKGQVKKIVDEL 227
DB 540 RYVPGATTTSSAATE-----TLENYKCKNFTSLIKLAGSGOSTETRAANKELVQNL 593
QY 228 LPAKIEAEETRKLYELKSPQPHLVPLKRSVALROLLPNSQSFIOQVQO-----TS 283
DB 594 LQKIEAEDFTSLYRELNSPPQVLYVPLKSLPALRLQTLTDSAAFIQSQSQQPPPTPS 653
QY 284 SDNVIACTTTTYSPTVTTVSSSQSEKSIYSGATAPRTVSVQTLNPLAGPVGAKGV 343
DB 654 Q-----ATYLTAVLVLSVQRTAGKTAATVTSALQPPVSL----- 690
QY 344 VTLHSVGPAAAGGTAGTGLQTSKPLVTSVANVTYTSILPEKPVSGTAVTSLPAPV 403
DB 691 -----TQPTQVGVGKQGPPLVYQ-----OPKP-----GALIRPPQV 724
QY 404 TREGTSGAICLPVSPVVSFCWDHICKPVGIGTVQIKLAGPVLSPAGIPTGSSSKQ 463
DB 725 TLQGT-----PMVALQPH-NRIMLTTPQIQI----- 751
QY 464 LPSLFHVQVQPSGGENKQVTTISHSSTLTQKCGQKTMPTVNTIIPISQPPASIIKQITL 523
DB 752 ----- 765
QY 524 PGKIL---SLOASPQKNRIKENVTSQPRDE 552
DB 766 PGTKALSAVSAQAAAQKNKLEPGGGSFRDD 797
RESULT 9
PCT-US02-30474-3271
Sequence 3271, Application PC/TUS0230474
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyun
APPLICANT: Xue, Aidong J.
APPLICANT: Zhao, Qiong A.
APPLICANT: Wang, Jian-Rui
APPLICANT: Wehrman, Tom
APPLICANT: Zhou, Ping
APPLICANT: Ghosh, Malabika
APPLICANT: Wang, Dunrui
APPLICANT: Ma, Yundong
APPLICANT: Asundi, Vinod
APPLICANT: Wang, Zhiwei
APPLICANT: Wang, Gezhi
APPLICANT: Haley-Vicente, Dana
APPLICANT: Drmanac, Radoje T
TITLE OF INVENTION: Novel Nucleic Acids and
FILE REFERENCE: Polypeptides
FILE REFERENCE: 810CIP PCT
CURRENT APPLICATION NUMBER: PCT/US02/30474
CURRENT FILING DATE: 2002-09-24
PRIOR APPLICATION NUMBER: US 60/324,631
PRIOR FILING DATE: 2001-09-24
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02623
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03


```
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28
; Remaining prior application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 3476
; SOFTWARE: PL_FL_genes Version 6.0
; SEQ ID NO 3271
; LENGTH: 1083
; TYPE: PRM
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1083)
; OTHER INFORMATION: Xaa = any amino acid or symbol as shown in the table 8 as set for
PCT-US02-30474-3271
```

Query Match 19.4%; Score 535; DB 1; Length 1083;

Best Local Similarity 29.1%; Pred. No. 1.1e-33; Matches 168; Conservative 65; Mismatches 145; Indels 200; Gaps 19;

```
QY 13 PPKVSSG-----PRLPAPQIVAVKAPNTTTFQFPANLQLPFGTVLIRKNSGPI 61
DB 414 PTATTSIGIRATLPPTVLAPRLPQP-----PQNTPTNIO---NFDLPDMVLVRENGCLL 464
QY 62 LVSPQOQIVTR-----AETTSNITSRPAPVPANPQTVKICIVPNSSSOLIKKVAATPYKKL 115
DB 465 MI-PQOALAQOQAQAHAPQPTTMAPRPATPISAPPVQISTQITGSGPLTR-RCPINII 522
QY 116 AQIGTIVTVTP-----KPSVSQVAVPTSVTVTP-----GKPLNTVTILKSSSL 161
DB 523 SRVSAQOTVQPSATLKRSPSQPOLVGLCAKATAXLGTAVQGTGPRV----- 575
QY 162 GASSTPSENEPLKAKENSAVOINLSPTMLENVKCKNFAMLIKACSGSPKMGONVK 221
DB 576 GATTTS-----VARETMENVKCKNFRSTLIKASSGKOSTETAAVVK 619
QY 222 KLVQLDARKEAEFTRKLYVELKSSPOPHLVFLKRSYVALRQLLPNSQSFIOGCVOQ 281
DB 620 ELVONLIDGKTEADFTSRILKRELNSSPOPLVFLKRSIPALQLPDSAAFIQSOQO 679
QY 282 ----TSSDMVIATCTTIVTSSPVVTVSSQSEKSIIVSGATAPRVSVQTLNPLAGPV 337
DB 680 PPPPTSQ-----ATTALFAVLSSSVQRTAGTAATVTSALQPVLSL----- 722
QY 338 GAKGVVTLHSGPTAGTGTAGTGLQTSKPLVTSVANTVTVTSIQPEPVVSGTAAT 397
DB 723 -----TQPTQVGVGQOGPTPLVIO-----QPKP-----GAL 750
QY 398 LSLPVTGETSGAATCLPSVKPVVSCMDHICKPVIGTVPQIKLAGVPLSQPAGIPT 457
DB 751 IRPQVLTQV-----PMVALRQPH-NRIMLTTPQIQIOL----- 783
QY 458 GSSSKQLFSLFHVVOQPSGNEKQVTTISHSSTLTIQKCGKTMVNTIPTISQPPASI 517
DB 784 -----NPLQPVV 791
QY 518 LKQITLPGNKTL---SIQASPTOKNRRIKENVTSQFRDE 552
DB 792 VKPAVLPGTKALSAVSAQAAAAAKNKLKEPGGSGFRD 829
```

RESULT 10
PCT-US01-20592-14

```
; Sequence 14, Application PC/TUS0120592
; GENERAL INFORMATION:
; APPLICANT: ANADYS PHARMACEUTICALS, INC.
; APPLICANT: THOMPSON, Craig
; APPLICANT: MOORE, Jeffrey
; APPLICANT: HURMAN, Ed T.
; APPLICANT: BRADLEY, John
```

```
; APPLICANT: DESIIVA, Thamara
; APPLICANT: HARRIS, Sandra
; APPLICANT: KOMARNITSKY, Svetlana
; APPLICANT: MENDILLO, Marc
; APPLICANT: MOORE, Daniel
; APPLICANT: MCCOY, Melissa
; APPLICANT: SANDERSON, Karen
; APPLICANT: HAO, Tariq
; APPLICANT: ZHU, Shuhao
; APPLICANT: LONG, Fan
; APPLICANT: DAVIDOV, Eugene
; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
; FILE REFERENCE: 0342/1G548-US1
; CURRENT APPLICATION NUMBER: PCT/US01/20592
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,164
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/224,457
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 1023
; TYPE: PRM
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: human genbank accession #: CAA72189
; NAME/KEY: misc_feature
; OTHER INFORMATION: Corresponds to SEQ ID NO: 87
PCT-US01-20592-14
```

Query Match 18.9%; Score 521; DB 1; Length 1023;

Best Local Similarity 31.7%; Pred. No. 1.4e-32; Matches 168; Conservative 73; Mismatches 157; Indels 132; Gaps 22;

```
QY 13 PPKVSSG-----PRLPAPQIVAVKAPNTTTFQFPANLQLPFGTVLIRKNSGPI 61
DB 414 PTATTSIGIRATLPPTVLAPRLPQP-----PQNTPTNIO---NFDLPDMVLVRENGCLL 464
QY 62 LVSPQOQIVTR-----AETTSNITSRPAPVPANPQTVKICIVPNSSSOLIKKVAATPYKKL 115
DB 465 MI-PQOALAQOQAQAHAPQPTTMAPRPATPISAPPVQISTQITGSGPLTR-QVTP----- 518
QY 116 AQIGTIVTVTPKPSVQVAVPTSVTVTPGKPLNTYT---TLKPSIGASS-----TPS 168
DB 519 ----TTIIKQV---SQOITVQPSHTLQRSQVQPOLVGLGAQOTASLGTAVQGTGPQ 571
QY 169 NE-PNLKAKENSAVOINLSPTMLENVKCKNFAMLIKACSGSPKMGONVKRYEOL 227
DB 572 RTVPGATTSSAAE-----TMENVKCKNFSLTIKILASSGKOSTETAAVKEIVONL 625
QY 228 LDAITEAEFTRKLYVELKSSPOPHLVFLKRSYVALRQLLPNSQSFIOGCVOQ-----TS 283
DB 626 LDGKTEADFTSRILKRELNSSPOPLVFLKRSIPALQLPDSAAFIQSOQOPPPPTS 685
QY 284 SDMVIATCTTIVTSSPVVTVSSQSEKSIIVSGATAPRVSVQTLNPLAGPVGAKGV 343
DB 686 Q-----ATTALFAVLSSSVQRTAGTAATVTSALQPVLSL----- 722
QY 344 VTLHSGVPTAATGTTAGTGLQTSKPLVTSVANTVTVTSIQPEPVVSGTAATVLSLPAV 403
DB 723 -----TQPTQVGVGQOGPTPLVIO-----QPKP-----GALIRPQV 756
QY 404 TFGETSGAATCLPSVKPVVSCMDHICKPVIGTVPQIKLAGVPLSQPAGIPTGS 459
DB 757 TLTQV-----PMVALRQPH-NRIMLTTPQIQVLSSESARILATNSELVGTILRS 804
QY 460 SSKQLFSLFHVVO-----OPSGNEKQ---VTTISHSSTLTIQKCGK 499
DB 805 CKDEFLLQAPLQRIKLEIGKKGITELHPDVSVISHAQOQLONLVEK 854
```

```

RESULT 11
PCT-US01-20592A-14
Sequence 14, Application PC/TUS0120592A
GENERAL INFORMATION:
APPLICANT: ANADYS PHARMACEUTICALS, INC.
APPLICANT: THOMPSON, Craig
APPLICANT: MOORE, Jeffrey
APPLICANT: BURMAN, Ed T.
APPLICANT: BRADLEY, John
APPLICANT: DESILVA, Thamara
APPLICANT: HARRIS, Sandra
APPLICANT: KOMARNITSKY, Svetlana
APPLICANT: MOORE, Daniel
APPLICANT: MCCOY, Melissa
APPLICANT: SANDERSON, Karen
APPLICANT: HAO, Tariq
APPLICANT: ZHU, Shuhao
APPLICANT: LONG, Fan
APPLICANT: DAVIDOV, Eugene
TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
FILE REFERENCE: 0342/16548-US1
CURRENT APPLICATION NUMBER: PCT/US01/20592A
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/215,164
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: US 60/224,457
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 146
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 1023
TYPE: PRT
ORGANISM: Homo sapiens
NAME/KEY: misc_feature
OTHER INFORMATION: Corresponds to SEQ ID NO: 87
PCT-US01-20592A-14

Query Match      18.9%; Score 521; DB 1; Length 1023;
Best Local Similarity 31.7%; Pred. No. 1,4e-32;
Matches 168; Conservative 73; Mismatches 157; Indels 132; Gaps 22;

QY 13 PRVSSG-----PRLRPQIVAVKAPRTTITIOFANILQPLPGYILINSNGPLM 61
DB 414 PIAATGIRATLTPVLPALPLP-----PQNPNTNQ--NQPLPGWLVSENGQL 464
QY 62 LVSPQOTVR-----AETTSNITSRPVAVPANPOTVKTICVPPSSSOLIKKVAVTVPK 115
DB 465 MI-PQOALQMOQAHQAPQTTMAPRPAPTTSAPVQISTVQAPGPIITAR-QVTP----- 518
QY 116 AOTGTTVTTVPKSSVQVAVPTSVVYTPGKPLNTVT--TLKPSLGLASS-----TPS 168
DB 519 ----TIIKQV---SOAQTVQPSATLQNSPGVQPOLVLGGAQOTSLGTAFAVQGTGP 571
QY 169 NE-PNLKAENSAVQINLSPTMLENKKCNFLAMLIKLAGSSQSPENKQNTKLVLEOL 227
DB 572 RIVPGATTTSSAATE-----TMENVKKCNFLSTLIKLAGSSQSTETAAVKELOVL 625
QY 228 LPAKIEAEFTRLKLYELKSSPQPHLVPLFKKSVALROLPLPSOSFIQOCVQO-----TS 283
DB 626 LQKIEADEFTSLRYELKSSPQPHLVPLFKKRLPALROLTPSNAFIQOSQOQPPPTTS 665
QY 284 SDNVIAVCTTGVTTSPVVTTVSSQSEKSIIVSGATAPRTVSQTLNPLAPGVAKAGV 343
DB 686 Q-----ATTALTAVVLSSSVQRTAGKTAATVTSALQPPVLSL-----722
QY 344 VTLHSVGPRAAGGTAGAGGLQTSKPLVTSVANTVTVSSLOPEKRVVSGTAVTSLP 403
DB 723 -----TQPIQVGVGKQGOQPPVLQ-----QPKP-----GALLIRPPQV 756
QY 404 TREGTSGAALCLPSVAVKFCMDHICKPVGTPVQIKLAQPGPVL-----SOPAGIPTGS 459

```

```

DB 757 TLQOT-----PMVALRQPH-NRIMLTTPQOVNLSSESRILATNSLVGLTTRS 804
QY 460 SSKQLESLFHVQ-----OPSGCNEKQ---VTTISHSSSTLTIQKQGR 499
DB 805 CKDETFLQAPLQRIELIKKHGITEHLPDVSVSHATQORLQNLVER 854

RESULT 12
US-09-893-519A-14
Sequence 14, Application US/09893519A
GENERAL INFORMATION:
APPLICANT: ANADYS PHARMACEUTICALS, INC.
APPLICANT: THOMPSON, Craig
APPLICANT: MOORE, Jeffrey
APPLICANT: BURMAN, Ed T.
APPLICANT: BRADLEY, John
APPLICANT: DESILVA, Thamara
APPLICANT: HARRIS, Sandra
APPLICANT: KOMARNITSKY, Svetlana
APPLICANT: MENDILLO, Marc
APPLICANT: MOORE, Daniel
APPLICANT: MCCOY, Melissa
APPLICANT: SANDERSON, Karen
APPLICANT: HAO, Tariq
APPLICANT: ZHU, Shuhao
APPLICANT: LONG, Fan
APPLICANT: DAVIDOV, Eugene
TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
FILE REFERENCE: 0342/16548-US2
CURRENT APPLICATION NUMBER: US/09/893,519A
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/215,164
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: US 60/224,457
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 146
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 1023
TYPE: PRT
ORGANISM: Homo sapiens
NAME/KEY: misc_feature
OTHER INFORMATION: Corresponds to SEQ ID NO: 87
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Human Genbank/CA472189
DATABASE ENTRY DATE: 1997-06-25
RELEVANT RESIDUES: (1)..(1023)
US-09-893-519A-14

Query Match      18.9%; Score 521; DB 22; Length 1023;
Best Local Similarity 31.7%; Pred. No. 1,4e-32;
Matches 168; Conservative 73; Mismatches 157; Indels 132; Gaps 22;

QY 13 PRVSSG-----PRLRPQIVAVKAPRTTITIOFANILQPLPGYILINSNGPLM 61
DB 414 PIAATGIRATLTPVLPALPLP-----PQNPNTNQ--NQPLPGWLVSENGQL 464
QY 62 LVSPQOTVR-----AETTSNITSRPVAVPANPOTVKTICVPPSSSOLIKKVAVTVPK 115
DB 465 MI-PQOALQMOQAHQAPQTTMAPRPAPTTSAPVQISTVQAPGPIITAR-QVTP----- 518
QY 116 AOTGTTVTTVPKSSVQVAVPTSVVYTPGKPLNTVT--TLKPSLGLASS-----TPS 168
DB 519 ----TIIKQV---SOAQTVQPSATLQNSPGVQPOLVLGGAQOTSLGTAFAVQGTGP 571
QY 169 NE-PNLKAENSAVQINLSPTMLENKKCNFLAMLIKLAGSSQSPENKQNTKLVLEOL 227
DB 572 RIVPGATTTSSAATE-----TMENVKKCNFLSTLIKLAGSSQSTETAAVKELOVL 625
QY 228 LPAKIEAEFTRLKLYELKSSPQPHLVPLFKKSVALROLPLPSOSFIQOCVQO-----TS 283
DB 626 LQKIEADEFTSLRYELKSSPQPHLVPLFKKRLPALROLTPSNAFIQOSQOQPPPTTS 685

```



```
RESULT 15
; US-60-173-464-20611
; Sequence 20611, Application US/60173464
; GENERAL INFORMATION:
; APPLICANT: LI, Peter W.D.
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000173
; CURRENT APPLICATION NUMBER: US/60/173,464
; CURRENT FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 30269
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20611
; LENGTH: 921
; TYPE: PRF
; ORGANISM: Drosophila
; US-60-173-464-20611

Query Match          9.4%; Score 258; DB 27; Length 921;
Best Local Similarity 22.9%; Pred. No. 3,6e-11;
Matches 135; Conservative 86; Mismatches 193; Indels 176; Gaps 24;

QY 25 POTVAKAPNTTIOIPANQLPPGVILKSNGLML-----VSPQOTVRAETTS 76
DB 133 POSPSITLSTLNTGTPA-----LVKTDNGFQLRVGTTGPTTVTQTITNTSNNS 184
QY 77 NITSRAVPANPQTVKICVTPNSSQ-----LIKVAVTPVKLAIGTTVVTVTP 127
DB 185 NTSSTNNHPTTQ-IRLQVPAASMTNTATSNITIVNSVASSGYANSSQPHLTQLMAQ 243
QY 128 KPSVSQSAVPSTVTVTPKPLNTVTTLKPSLSGASSTPSNEPNLKAENSAVQINLSP 187
DB 244 APQLPQITQITIPAOOSQOQVNNVSSAGTFAVSSSTA-----ATT 287
QY 188 TMLENYK-KCKNFLAMLIKACSGSOSPENQOVKKIQLDPAKTEAEETKRLVYELK 246
DB 288 TOGNTKCKCRKLANLIEL--STREPKYKAVKRLIQLVNAVNPPEFCRDLERLN 345
QY 247 SSBQPHLVPLKKSVALRQL-----LPNSQSFQ--- 276
DB 346 ASPOPCLLIGLKKSLPLRLQALYTKELVIEGKPPQHVHGLAGLSQQLPKIQAIIRPIG 405
QY 277 ----QCVQOTSSDMVATCTTVTTSPTVTVTVSSSQSEKSIIVSGATAPRTVS---VOT 329
DB 406 PSQTTTIGQTVQVMI--TPNALGTTPRTTIGHTTISKOPN--IRLPTARLVNTGIGIRT 460
QY 330 LNPFLAGVGAAGVATLHNSVPTAATGTTAGTGLQTSPLVTSVANTVTTVSLQPEKP 389
DB 461 QTP-SLQVPGQANIVQIR--GRQHAQLQRTGSQVIRATIRP-----PNSVPTAN----- 506
QY 390 VVSGTAVTLSLPAVTEGETSGAAILCPYKPVVSVFCWDHICKPVGIGTPVOIKLAQPGVL 449
DB 507 -----KLFAVKVGGTQIKAI-TPSLHP-----PSL 530
QY 450 SQPAGITPTGSSSKQLSLPHVVOQPSGNGEKQVTTTISHSSTLTLOKCGQKTMPTVNTIIP 509
DB 531 AATSGGPPPTPLTSLVS-----TLNSAST-----TTLPIPS-LPT 564
QY 510 SQFPFASI-----LKQITLPGN-----KILSLQA---SPTOKRIKENVT 546
DB 565 VHLPPALAREQMOMSLNHNHNPDAKLVEIKAPSLHPPHMERINASLT 614
```

Search completed: February 16, 2003, 22:18:18
Job time : 204.796 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 22:00:19 ; Search time 22.4582 Seconds

(without alignments)
2076.603 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_552

Perfect score: 2758
Sequence: 1 GLVTRVAPVSAAPKVSQGP.....ASPTQKNRIKENYSCPRDE 552

Scoring table: BLASTSUM62
Gapop 10.0 , Gapext 0.5

Searched: 429898 seqs, 84487048 residues

Total number of hits satisfying chosen parameters: 429898

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	539	19.5	709	US-09-724-676-63927	Sequence 63927, A
2	539	19.5	709	US-09-724-676-63927	Sequence 63927, A
3	539	19.5	741	US-09-724-676-63930	Sequence 63930, A
4	539	19.5	741	US-09-724-676-63930	Sequence 63930, A
5	539	19.5	805	US-09-724-676-63921	Sequence 63921, A
6	539	19.5	805	US-09-724-676-63921	Sequence 63921, A
7	539	19.5	837	US-09-724-676-63924	Sequence 63924, A
8	539	19.5	837	US-09-724-676-63924	Sequence 63924, A
9	527	19.1	925	US-09-724-676-63928	Sequence 63928, A
10	527	19.1	925	US-09-724-676-63928	Sequence 63928, A
11	527	19.1	957	US-09-724-676-63931	Sequence 63931, A
12	527	19.1	957	US-09-724-676-63931	Sequence 63931, A
13	527	19.1	1021	US-09-724-676-63922	Sequence 63922, A
14	527	19.1	1021	US-09-724-676-63922	Sequence 63922, A
15	527	19.1	1053	US-09-724-676-63925	Sequence 63925, A
16	527	19.1	1053	US-09-724-676-63925	Sequence 63925, A
17	519	18.8	679	US-09-724-676-63929	Sequence 63929, A
18	519	18.8	679	US-09-724-676-63929	Sequence 63929, A
19	519	18.8	711	US-09-724-676-63920	Sequence 63920, A
20	519	18.8	711	US-09-724-676-63920	Sequence 63920, A
21	519	18.8	775	US-09-724-676-63923	Sequence 63923, A
22	519	18.8	775	US-09-724-676-63923	Sequence 63923, A
23	519	18.8	807	US-09-724-676-63926	Sequence 63926, A
24	519	18.8	807	US-09-724-676-63926	Sequence 63926, A
25	219	7.9	5179	US-60-438-735-151	Sequence 151, App
26	206.5	7.5	1322	US-10-218-140-4828	Sequence 4828, App

27	205.5	7.5	2108	1	PCT-US02-33723-2	Sequence 2, Appli
28	205.5	7.5	2135	6	US-10-288-798-9	Sequence 9, Appli
29	205.5	7.5	2382	6	US-10-196-935A-2	Sequence 2, Appli
30	205.5	7.5	2382	6	US-10-293-017-48	Sequence 48, Appli
31	205.5	7.5	2382	6	US-10-052-648A-40	Sequence 40, Appli
32	205.5	7.5	2382	6	US-10-293-071-48	Sequence 48, Appli
33	201.5	7.3	5374	6	US-10-028-248A-75	Sequence 75, Appli
34	201.5	7.3	5374	6	US-10-107-782-75	Sequence 75, Appli
35	199.5	7.2	2601	6	US-10-028-248A-76	Sequence 76, Appli
36	199.5	7.2	2601	6	US-10-107-782-76	Sequence 76, Appli
37	197.5	7.2	5376	6	US-10-028-248A-74	Sequence 74, Appli
38	197.5	7.2	5376	6	US-10-107-782-74	Sequence 74, Appli
39	194.5	7.1	2781	6	US-10-263-929-122	Sequence 122, Appli
40	194	7.0	2971	6	US-10-218-140-1990	Sequence 1990, App
41	192.5	7.0	1794	7	US-60-427-045-299	Sequence 299, App
42	192.5	7.0	1799	7	US-60-427-045-149	Sequence 149, App
43	192.5	7.0	1821	7	US-60-427-045-82	Sequence 82, Appli
44	192.5	7.0	11721	7	US-60-427-045-162	Sequence 162, App
45	190.5	6.9	5935	6	US-10-243-243A-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-09-724-676-63927
: Sequence 63927, Application US/09724676
: GENERAL INFORMATION:
: APPLICANT: CompuGen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 CompuGen
: CURRENT APPLICATION NUMBER: US/09/724, 676
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 63927
: LENGTH: 709
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-724-676-63927

Query Match 19.5%, Score 539, DB 5, Length 709;
Best Local Similarity 30.2%, Pred. No. 3, 1e-22;
Matches 172; Conservative 66; Mismatches 144; Indels 188; Gaps 21;

QY	13	PKVSSG-----PRLPAPQIVAKAANTTITQPPNALQLPFGTVLIKNSGRLM 61
DB	286	PTATTSGIRATLTPVLAIRLPQ-----PONPTNIO---NFQLPQGVILVRSNGQL 336
QY	62	LVSPOQVTR-----AETTSNITSRPAPVAPQVFKICTVYNSSSOLIKKVAVPPVKL 115
DB	337	MI-PQALAQMOQAHAQPPQTMARPRATPPISAPVQISTVQAPPTIAR-QVTP----- 390
QY	116	AOIGTTVTVTPKSSVSVAVPTISVTVTPGKPLNTVT--TLKPSLCASS-----TPS 168
DB	391	---TTIIKQV---SQAQTVQPSATLQRPQVQQLVIGGAQAASIGTAAVQTGPQ 443
QY	169	NE-PLKENSAAVQINISPTLENVKCKKNTLAILIACSGSSPEMGQVAKVLVEOL 227
DB	444	RTVPATTTSSAATE-----TMENVKCKKNTLILIKASSGKSTFAANVKELVONTL 497
QY	228	IDAKTEAEETPKRLVVELKSSPOPHLVPFLKSSVALLPNSQSFIOQCVO-----TS 283
DB	498	LDGKTEADPTSRKLYRELNSSQPLVTFPKLSLALQOLTPDASAFIOSSQOQPPPTPS 557
QY	284	SDMVIATCTTTVTSSPVVTTTVSSQSSEKSIIVSGATAPRVSVQTLNPLAGPVGAKGV 343
DB	558	Q-----ATTALTFVAVLSSSVQRTAGKTAATVTSALQPVLSL----- 594
QY	344	VLHSGVPTAATGCTAGTGLQTSKPLVTSVANTVTVSLQPEKPVVSGRAVITSLP 403
DB	595	-----TQPVQGVGKQGPPLVIO-----QPKP-----GALIRPQV 628

```
OY 404 TFGTSGAATCLPSVKPVSEFCMDHICKPVIGTPVOIKLAQPGPVLSQAPAGIPTGSSSQ 463
DB 629 TLTOT-----PMVALRQPH-NRIMLTTPQOQL----- 655
OY 464 LFSLFHVVOQPSGSGNEKOVYTTISHSSTLTIOKCGOKTMTPTIIPISQFPASILKOITL 523
DB 656 -----NPLQPVVVKRAVIL----- 669
OY 524 PGKNTL---SLQASPTOKNRKIKENVTSCFR 550
DB 670 PGTKALSAVSAQAAAAQKNKLEKPEGGSFR 659

RESULT 2
US-09-724-676A-63927
; Sequence 63927, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63927
; LENGTH: 709
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-63927

Query Match 19.5%; Score 539; DB 5; Length 709;
Best Local Similarity 30.2%; Pred. No. 3.1e-22;
Matches 172; Conservative 66; Mismatches 144; Indels 188; Gaps 21;

OY 13 PKVSSG-----PRLPAPQIVAVKAPNTTIOFPANLQPLPGTVLIXNSGSLM 61
DB 286 PRATSGIRATLPTVLAAPRLPOP-----PONPTNIQ---NQPLPGVAVLHSENGQL 336
OY 62 LVSPQQTVTR-----AETTSNITSRPAVNPQTVKICIVPNSSSOLIKKAVATPVKKL 115
DB 337 MI-PQALAQMOQAHAQPOPTMAPRAPPTSAAPVOISTVOAGCTPIIAR-QVTP----- 390
OY 116 AQIGTVVTVTPKPSVQSAVAVPTSVTVTPGKPLNTVT--TLKPSLGSAS-----TPS 168
DB 391 -----TLTIKQV---SQAQITVOPSATLORSPGVQPOLVLGAAQTASLGTATAVQGTGP 443
OY 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFKLAMLKILACSGSOSPEMGONVKKILEOL 227
DB 444 RYVPGATTTSSATE-----TMENVKCKNFKLSTILKILASGKOSTETANVKELVQNL 497
OY 228 LDKAIEAEFTRLKLYELKSSPOHLPFLKKSVALROLPLNSOSFTIOQVOO-----TS 283
DB 498 LDKAIEAEFTSLRLYRELNSPOHLPFLKRLPALROLPLPSAARTIOOSQOQPPPTPS 557
OY 284 SDVVIATCTTIVTSPVTTVSSOSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGV 343
DB 558 Q-----ATTALTAVLVSSSVQRTAGKTAAYTSALQPVLSL----- 594
OY 344 VTLHVGPTAATGTTAGTGLQTSKPLVTVSVANTVTVTSLOPEKPVVSGTAVTSLPAPV 403
DB 595 -----TQPTQVGVKQOGQPTPLVIO-----QPKP-----GALIRPPQV 628
OY 404 TFGTSGAATCLPSVKPVSEFCMDHICKPVIGTPVOIKLAQPGPVLSQAPAGIPTGSSSQ 463
DB 629 TLTOT-----PMVALRQPH-NRIMLTTPQOQL----- 655
OY 464 LFSLFHVVOQPSGSGNEKOVYTTISHSSTLTIOKCGOKTMTPTIIPISQFPASILKOITL 523
DB 656 -----NPLQPVVVKRAVIL----- 669
OY 524 PGKNTL---SLQASPTOKNRKIKENVTSCFR 550
DB 670 PGTKALSAVSAQAAAAQKNKLEKPEGGSFR 659
```

```
RESULT 3
US-09-724-676-63930
; Sequence 63930, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63930
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-63930

Query Match 19.5%; Score 539; DB 5; Length 741;
Best Local Similarity 30.2%; Pred. No. 3.2e-22;
Matches 172; Conservative 66; Mismatches 144; Indels 188; Gaps 21;

OY 13 PKVSSG-----PRLPAPQIVAVKAPNTTIOFPANLQPLPGTVLIXNSGSLM 61
DB 318 PRATSGIRATLPTVLAAPRLPOP-----PONPTNIQ---NQPLPGVAVLHSENGQL 368
OY 62 LVSPQQTVTR-----AETTSNITSRPAVNPQTVKICIVPNSSSOLIKKAVATPVKKL 115
DB 369 MI-PQALAQMOQAHAQPOPTMAPRAPPTSAAPVOISTVOAGCTPIIAR-QVTP----- 422
OY 116 AQIGTVVTVTPKPSVQSAVAVPTSVTVTPGKPLNTVT--TLKPSLGSAS-----TPS 168
DB 423 -----TLTIKQV---SQAQITVOPSATLORSPGVQPOLVLGAAQTASLGTATAVQGTGP 475
OY 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFKLAMLKILACSGSOSPEMGONVKKILEOL 227
DB 476 RYVPGATTTSSATE-----TMENVKCKNFKLSTILKILASGKOSTETANVKELVQNL 529
OY 228 LDKAIEAEFTRLKLYELKSSPOHLPFLKKSVALROLPLNSOSFTIOQVOO-----TS 283
DB 530 LDKAIEAEFTSLRLYRELNSPOHLPFLKRLPALROLPLPSAARTIOOSQOQPPPTPS 589
OY 284 SDVVIATCTTIVTSPVTTVSSOSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGV 343
DB 590 Q-----ATTALTAVLVSSSVQRTAGKTAAYTSALQPVLSL----- 626
OY 344 VTLHVGPTAATGTTAGTGLQTSKPLVTVSVANTVTVTSLOPEKPVVSGTAVTSLPAPV 403
DB 627 -----TQPTQVGVKQOGQPTPLVIO-----QPKP-----GALIRPPQV 660
OY 404 TFGTSGAATCLPSVKPVSEFCMDHICKPVIGTPVOIKLAQPGPVLSQAPAGIPTGSSSQ 463
DB 661 TLTOT-----PMVALRQPH-NRIMLTTPQOQL----- 667
OY 464 LFSLFHVVOQPSGSGNEKOVYTTISHSSTLTIOKCGOKTMTPTIIPISQFPASILKOITL 523
DB 688 -----NPLQPVVVKRAVIL----- 701
OY 524 PGKNTL---SLQASPTOKNRKIKENVTSCFR 550
DB 702 PGTKALSAVSAQAAAAQKNKLEKPEGGSFR 731

RESULT 4
US-09-724-676A-63930
; Sequence 63930, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
```

; NUMBER OF SEQ ID NOS: 97222
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 63930
 ; LENGTH: 741
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-724-676A-63930

Query Match 19.5%; Score 539; DB 5; Length 741;
 Best Local Similarity 30.2%; Pred. No. 3.2e-22;
 Matches 172; Conservative 66; Mismatches 144; Indels 188; Gaps 21;

QY 13 PPKVSSG-----PRUPAIOYAVKAPNTTTIOFPANLQIPGTVLIRKSNCGPLM 61
 Db 318 PTATSGIRATLTPTVLAPRLPQ-----PQNPNTNIO--NFQLPQPMVLRSENGQL 368
 QY 62 LVSPQOQVYTR-----AETTSNITSRPVAVPANPQTVKICVPSNSSOLIKKAVATPVKKL 115
 Db 369 MI-PQOALAQOQAQAHAPQPTTMARPRATPTPSAPPVQISTVOABQPTLIAR-QVTP----- 422
 QY 116 AOKTAEETFRKLYVELKSSPQPHLYPFLKSSVVALRQLLPNSQSFIOQCVQO-----TS 168
 Db 423 -----TTIIKQV-----SQOQTVQPSATLQSRPGVOPQVLGGAQOASLGTAFAVOTGTIPQ 475
 QY 169 NE-PNLKAENSAVAOINLSPMLENVKCKNFMLMLIKLACSGSQSPMGONVKKLYBQL 227
 Db 476 RTVPGATTTSSAATE-----TMENVKCKNFSLTLIKLACSGSQSTETANVKELVQNL 529
 QY 228 LDKAIEAEETFRKLYVELKSSPQPHLYPFLKSSVVALRQLLPNSQSFIOQCVQO-----TS 283
 Db 530 LDGKIEADEFTSRILYRELNSSPQPLVPEFLKRLSLPALRQLPDSAAFIQSSQOQPPPTPS 589
 QY 284 SDMYATCTTTVTSPPVTTVSSSQSEKSIIVSGATAPRTVSVQTLNPLAGPVGAKAGV 343
 Db 590 Q-----ATTALNAVVLSSVQRTAGKTAATVTSALQPPVLSL----- 626
 QY 344 VTLHSGPPTAATGCTTACTGLQTSKPLVTSVANTVTVSSIQPEKPVVSGTAVTLSLPAV 403
 Db 627 -----TOPVOVGKQGOPTPLVIO-----QPPRP-----GALIRPPQV 660
 QY 404 TFGETSGAICLPVSKPVVSEFCMDHICKPVIGTPOVQIKLAPGPVLSQAPAGIPGSSSKQ 463
 Db 661 TLTQTP-----PWALRQPH-NRIMLTTPQOIOI----- 687
 QY 464 LFSLEHVQOPSGGNEKQVTTIHSSTLTIOKCGKMTPVNTIIPTSQPPASILKQITL 523
 Db 688 -----NLOQVVPVVKRAVL 701
 QY 524 PGNKIL---SIQASPTQKNRIKENVTSCFR 550
 Db 702 PGTALSAVSAQAAAAQKNKLEKPEGGSFR 731

RESULT 5
 US-09-724-676-63921
 ; Sequence 63921, Application US/09724676
 ; GENERAL INFORMATION:
 ; APPLICANT: Compugen LTD
 ; TITLE OF INVENTION: Variants of alternative splicing
 ; FILE REFERENCE: 129181.4 Compugen
 ; CURRENT APPLICATION NUMBER: US/09/724.676
 ; CURRENT FILING DATE: 2000-11-28
 ; NUMBER OF SEQ ID NOS: 97222
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 63921
 ; LENGTH: 805
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-724-676-63921

Query Match 19.5%; Score 539; DB 5; Length 805;
 Best Local Similarity 30.2%; Pred. No. 3.5e-22;
 Matches 172; Conservative 66; Mismatches 144; Indels 188; Gaps 21;

QY 13 PPKVSSG-----PRUPAIOYAVKAPNTTTIOFPANLQIPGTVLIRKSNCGPLM 61
 Db 382 PTATSGIRATLTPTVLAPRLPQ-----PQNPNTNIO--NFQLPQPMVLRSENGQL 432
 QY 62 LVSPQOQVYTR-----AETTSNITSRPVAVPANPQTVKICVPSNSSOLIKKAVATPVKKL 115
 Db 433 MI-PQOALAQOQAQAHAPQPTTMARPRATPTPSAPPVQISTVOABQPTLIAR-QVTP----- 486
 QY 116 AOKTAEETFRKLYVELKSSPQPHLYPFLKSSVVALRQLLPNSQSFIOQCVQO-----TS 168
 Db 487 -----TTIIKQV-----SQOQTVQPSATLQSRPGVOPQVLGGAQOASLGTAFAVOTGTIPQ 539
 QY 169 NE-PNLKAENSAVAOINLSPMLENVKCKNFMLMLIKLACSGSQSPMGONVKKLYBQL 227
 Db 540 RTVPGATTTSSAATE-----TMENVKCKNFSLTLIKLACSGSQSTETANVKELVQNL 593
 QY 228 LDKAIEAEETFRKLYVELKSSPQPHLYPFLKSSVVALRQLLPNSQSFIOQCVQO-----TS 283
 Db 594 LDGKIEADEFTSRILYRELNSSPQPLVPEFLKRLSLPALRQLPDSAAFIQSSQOQPPPTPS 653
 QY 284 SDMYATCTTTVTSPPVTTVSSSQSEKSIIVSGATAPRTVSVQTLNPLAGPVGAKAGV 343
 Db 654 Q-----ATTALNAVVLSSVQRTAGKTAATVTSALQPPVLSL----- 690
 QY 344 VTLHSGPPTAATGCTTACTGLQTSKPLVTSVANTVTVSSIQPEKPVVSGTAVTLSLPAV 403
 Db 691 -----TOPVOVGKQGOPTPLVIO-----QPPRP-----GALIRPPQV 724
 QY 404 TFGETSGAICLPVSKPVVSEFCMDHICKPVIGTPOVQIKLAPGPVLSQAPAGIPGSSSKQ 463
 Db 725 TLTQTP-----PWALRQPH-NRIMLTTPQOIOI----- 751
 QY 464 LFSLEHVQOPSGGNEKQVTTIHSSTLTIOKCGKMTPVNTIIPTSQPPASILKQITL 523
 Db 752 -----NLOQVVPVVKRAVL 765
 QY 524 PGNKIL---SIQASPTQKNRIKENVTSCFR 550
 Db 766 PGTALSAVSAQAAAAQKNKLEKPEGGSFR 795

RESULT 6
 US-09-724-676A-63921
 ; Sequence 63921, Application US/09724676A
 ; GENERAL INFORMATION:
 ; APPLICANT: Compugen LTD
 ; TITLE OF INVENTION: Variants of alternative splicing
 ; FILE REFERENCE: 129181.4 Compugen
 ; CURRENT APPLICATION NUMBER: US/09/724.676A
 ; CURRENT FILING DATE: 2000-11-28
 ; NUMBER OF SEQ ID NOS: 97222
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 63921
 ; LENGTH: 805
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-724-676A-63921

Query Match 19.5%; Score 539; DB 5; Length 805;
 Best Local Similarity 30.2%; Pred. No. 3.5e-22;
 Matches 172; Conservative 66; Mismatches 144; Indels 188; Gaps 21;

QY 13 PPKVSSG-----PRUPAIOYAVKAPNTTTIOFPANLQIPGTVLIRKSNCGPLM 61
 Db 382 PTATSGIRATLTPTVLAPRLPQ-----PQNPNTNIO--NFQLPQPMVLRSENGQL 432
 QY 62 LVSPQOQVYTR-----AETTSNITSRPVAVPANPQTVKICVPSNSSOLIKKAVATPVKKL 115
 Db 433 MI-PQOALAQOQAQAHAPQPTTMARPRATPTPSAPPVQISTVOABQPTLIAR-QVTP----- 486
 QY 116 AOKTAEETFRKLYVELKSSPQPHLYPFLKSSVVALRQLLPNSQSFIOQCVQO-----TS 168

```

Db 487 ----TTIIIGV---SQAGTTVPASATLQSPGVPQVLTGLGAAGTASLGTAVGCTPG 539
OY 169 NE-PULKENSAAYQINLSPTMLENKKCKKNFLAMLIKILACSSGSPSEMGQANKKIVLEOL 227
Db 540 RTVPQATTTSSAATE-----TWENYKCKKNFLSTLIKILASSCKOSTETPAANKKELVQNL 593
OY 228 LDKATIEAEFTTKLYVELKSSPOPHLVPLKKSVALROLPLPNSQSFIOQCVQO-----TS 283
Db 594 LDGKTEADPEFTSRILRELNSPQPLVPLPLKSLPALROLTPDSNAFLIQSSQOQPPPPPS 653
OY 284 SSMVATCTTYYTTSPPVYTTTSSSSSEKSIISGATAPRTVYSVOTLNLPLACPVGAKAGY 343
Db 654 Q-----ATTALTLVNLSSSVORTAGKTAAYTSALOPPVLSL----- 690
OY 344 VTLHVGPTAAAGGTAGGLQTSKPLVTSVANVTVTSLOPEKPVSGTAVTSLSPAV 403
Db 691 -----TQPPVQVGVGKQGOPTPLVYIQ-----QPPK-----GALLPPOV 724
OY 404 TFGFTSGAAILCPVSKPVVSCFMDHICKPVGIGTPVOQIKLAPGQPVLSOPAGIPTGSSSKO 463
Db 725 TLTOT-----PVVALRQPH-NRIMLTTPQOIQ----- 751
OY 464 LESLPHVYQPSGGENKQVTTTISHSSTLTIQCGQKTMVNTLITPSOPPASILKQITL 523
Db 752 -----NPLQPPVVPVAPVL 765
OY 524 PGNKIL---SLQASPTQKNRIKENVTSCPR 550
Db 766 PGTKALSAVSAQAAAAQAQKNKLTKEPGGSPR 795
RESULT 7
US-09-724-676-63924
; Sequence 63924, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Comugen Ltd
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Comugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63924
; LENGTH: 837
; TYPE: prt
; ORGANISM: Homo sapiens
; US-09-724-676-63924

```

```

Query Match      19.5%; Score 539; DB 5; Length 837;
Best Local Similarity 30.2%; Pred. No. 3.7e-22;
Matches 172; Conservative 66; Mismatches 144; Indels 188; Gaps 21.

QY   13 PPKVSSG-----PRLPAQIVAKAANTTITQEPANILQPPGVILIKSNGSLM 61
    : ||| |||| | | | | | | | | | | | | | | | | | | | | | |
Db   414 PRATSGIRATLTPVLAARLPQ-----PONPNIIQ--NFQLPRGNVLRSNGQLL 464
    : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY   62 LVSPOQTATR-----AETTSNITSRPAPVANPQVTKICTVPNSSQLIKKVAVTPVKTL 115
    : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   465 MI-PQALALOMQAQAAHPQOTTMADRRAPTASAPVQISTVQAQGPILIAR-QVTF---- 518
    : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY   116 AIGITVVTVTPKPSVGSVANPTSVTVTVTPKPLNTYT--TLKBSGLASS-----TPS 168
    : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   519 ----TTTIKOV---SQAQITVQSPSATLRSPKVOPOLVGAQAQIATATAVQGTGG 571
    : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY   169 NE-PIUKAENSAAYOINLSPTMLENVKKCKNLAMLIKLAGSGSOSPEKGONVKKLEVDL 227
    : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   572 RTVEPATITSSAATE-----IMENVKKCKNFLSTILIKLASSGKOSTETAAWVKELVOL 625
    : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY   228 LDAKIEAEFTFRKLVELKESBPQHILVFLKKSVALROLLENDSOFIQOCVQO----TS 283
    : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   626 LDGKIETAEADFTRSLRYRELNSSPQPLTVFLPKRSLPALRQLTDFDSMAFIQDSQQQPPPPIS 685
    : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY   284 SDMYATCTTTVTYSPPVTVTTVYSSSOSEKSIIVSGATAPRTVSYOTLNLAGPVGAKAGV 343

```

Dd	686	Q-----ATTALAAVAVLSSSVORTAGTATVYSAOPPLSL:-	722
Oy	344	VTLHSVGPTAATGGTTAGTGILLOTSKPLVTSVANTVTVLSOPKKPVSGTAATVLSLPAV	403
Dd	723	-----TPTLYGVGKGQGPFLVLQ-----QQPKP-----GALLRPQV	756
Oy	404	TTEGTSGAICLPYSVKPVVSCFCHICKIPVIGTPVOJIKLAOPGFVLISQAPGIFPGSSSKQ	463
Dd	757	TLTOR-:::--PWVALORPH-NKIMITTYOQTOL-----	793
Oy	464	LESLEHVVOQPDSGNGEKOVTVIISHSSTLIIOCGCKOTMPVNTIIPIVSQFPASILKOITL	523
Dd	784	-----NPLQPVVVVKA VL	797
Oy	524	PQNKL---SLQASPTOKNRILEKVTSQFR	550
Dd	798	PGRKRLSVNSQAQAAAQKNKLEKEGGGSFR	827

```

RESULT 8
US-09-724-676A-63924
: Sequence 63924, Application US/09724676A
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724.676A
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 63924
:
: LENGTH: 837
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-724-676A-63924

```

[illegible]

QY 464 LFSLEHVVOQPSGNGENKQVTTISHSSTLTIQCGQKTMPTVNTIITPSGPPASILKQITL 523
Db 784 -----NPLQPVVAVKPAVL 797
QY 524 GKNKIL---SLQASPTOKNRKIKENTSCFR 550
Db 798 GTRKLSAVSAQAQAAAKNKLKEPGGSFR 827

RESULT 9
US-09-724-676-63928

; Sequence 63928, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63928
; LENGTH: 925
; TYPE: PR1
; ORGANISM: Homo sapiens
US-09-724-676-63928

Query Match 19.1%; Score 527; DB 5; Length 925;
Best Local Similarity 30.6%; Pred. No. 1.9e-21;

Matches 175; Conservative 58; Mismatches 122; Indels 216; Gaps 24;

QY 13 PPKVSSG-----PRLPAPQIVAVKAPNTTIOFPANLQLPCTVLIKNSGPI 61
Db 286 PRTATSGRATLPTVLAAPRLPQ-----PQNTNIO---NFQLPQAMVLRSENGQL 336
QY 62 LVSPQQTIVTR-----AETTSNITSRAVPANPQTVKICVTPNSSSOLIKKVAATPVYK 115
Db 337 MI-PQQAIAQWQAQAHQAPQTTMAPRPTSPAPVOISTVQAPCTPIAR-QVTP----- 390
QY 116 AQTITVVTTPKSSVQSAVAVPTSVTVTGKPLNTVT--TLKPSLSGASS-----TPS 168
Db 391 -----TTIIKQV---SQAQTTVQPSATLQRSQVQPOLVLSGAQTAISLGTATVOTGPQ 443
QY 169 NE-PNLKAENSAVAQINLSPTMLENVKCKNFMLIKLACSGSOSPDMGONVKKIVEQL 227
Db 444 RTVGATTTSSAATE-----TMENVKCKNFSTLIKILASSGQSTETANVKELYONL 497
QY 228 LDKAIEEFTRLKLYELKSSPOPHLVPLFKSVVALRQLLPNSQSTIQCVQOTSSDMV 287
Db 498 LDKAIEEFTSRLYRELNSSPOPYLVPLFKRSLPALROLPPDSAAFTQOSQOQPPP-- 555
QY 288 IATCTTIVTSPVTTVSSSSEKSIIVSGATAPRTVSVOGTLNPLAGPVGAKAGVYTLH 347
Db 556 -----TSQATTA--LTAAVLS-----SVQ-----R 574
QY 348 SVGPPTAATGCTTAACTGLQTSKPLTVSVANTVTVTSLOPEKPVVSGTAVTLSPAVTGE 407
Db 575 TAGKTAAT-----VTSALQP--PVLSLT----- 595
QY 408 TSGAICLPVSPKPVVSCWDHICKPVIGTPVQI---KLAQGPVLSQAPAGIPTGSSSKOL 464
Db 596 -----VIOQP-----QPTQVGVKQGGQPTPL----- 611
QY 465 FSLFHVVOQPSGNGENKQVTTISHSSTLTIQCGQKTMPTVNTIITPSGPPASILKQITL 524
Db 612 -----VIOQP-----PKPG-----ALIQNLQLPVAVKPAVL 640
QY 525 GKNKIL---SLQASPTOKNRKIKENTSCFRDE 552
Db 641 GTRKLSAVSAQAQAAAKNKLKEPGGSFRD 671

RESULT 10
US-09-724-676A-63928

; Sequence 63928, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63928
; LENGTH: 925
; TYPE: PR1
; ORGANISM: Homo sapiens
US-09-724-676A-63928

Query Match 19.1%; Score 527; DB 5; Length 925;
Best Local Similarity 30.6%; Pred. No. 1.9e-21;

Matches 175; Conservative 58; Mismatches 122; Indels 216; Gaps 24;

QY 13 PPKVSSG-----PRLPAPQIVAVKAPNTTIOFPANLQLPCTVLIKNSGPI 61
Db 286 PRTATSGRATLPTVLAAPRLPQ-----PQNTNIO---NFQLPQAMVLRSENGQL 336
QY 62 LVSPQQTIVTR-----AETTSNITSRAVPANPQTVKICVTPNSSSOLIKKVAATPVYK 115
Db 337 MI-PQQAIAQWQAQAHQAPQTTMAPRPTSPAPVOISTVQAPCTPIAR-QVTP----- 390
QY 116 AQTITVVTTPKSSVQSAVAVPTSVTVTGKPLNTVT--TLKPSLSGASS-----TPS 168
Db 391 -----TTIIKQV---SQAQTTVQPSATLQRSQVQPOLVLSGAQTAISLGTATVOTGPQ 443
QY 169 NE-PNLKAENSAVAQINLSPTMLENVKCKNFMLIKLACSGSOSPDMGONVKKIVEQL 227
Db 444 RTVGATTTSSAATE-----TMENVKCKNFSTLIKILASSGQSTETANVKELYONL 497
QY 228 LDKAIEEFTRLKLYELKSSPOPHLVPLFKSVVALRQLLPNSQSTIQCVQOTSSDMV 287
Db 498 LDKAIEEFTSRLYRELNSSPOPYLVPLFKRSLPALROLPPDSAAFTQOSQOQPPP-- 555
QY 288 IATCTTIVTSPVTTVSSSSEKSIIVSGATAPRTVSVOGTLNPLAGPVGAKAGVYTLH 347
Db 556 -----TSQATTA--LTAAVLS-----SVQ-----R 574
QY 348 SVGPPTAATGCTTAACTGLQTSKPLTVSVANTVTVTSLOPEKPVVSGTAVTLSPAVTGE 407
Db 575 TAGKTAAT-----VTSALQP--PVLSLT----- 595
QY 408 TSGAICLPVSPKPVVSCWDHICKPVIGTPVQI---KLAQGPVLSQAPAGIPTGSSSKOL 464
Db 596 -----VIOQP-----QPTQVGVKQGGQPTPL----- 611
QY 465 FSLFHVVOQPSGNGENKQVTTISHSSTLTIQCGQKTMPTVNTIITPSGPPASILKQITL 524
Db 612 -----VIOQP-----PKPG-----ALIQNLQLPVAVKPAVL 640
QY 525 GKNKIL---SLQASPTOKNRKIKENTSCFRDE 552
Db 641 GTRKLSAVSAQAQAAAKNKLKEPGGSFRD 671

RESULT 11
US-09-724-676-63931
; Sequence 63931, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63931
; LENGTH: 957

Db 540 RTVPGATTSSAATE-----TMENVKCKNFSLTLIKLASSGKOSTETANVKELVONL 593
Qy 228 LDKAKIEEFTFKRLKVELKSSPOPLVPLFKKSVVALRQLLPNSOSFIQOCVOQTSSIMV 287
Db 594 LDGKIEADEFTSKRLRELNSSPOPLVPLFKKSLPALRQLLPDSAAFIQOQQOQPPPP-- 651
Qy 288 IATCTTATTSPPVTTTSSGSEKSIIVSGATAPRTVSVOTLNLAPGVGAKAGAVTLH 347
Db 652 ----TSQATTA--LTAVALSS-----SVQ-----R 670
Qy 348 SVGPAAAGCTTAGGLQTSKPLVTSVANTVTYVLSLOPEKPVVSGTAVTSLPAVTEGE 407
Db 671 TAGKTAAT-----VTSALP--PVLSLT----- 691
Qy 408 TSGAALCLPSVKPVVSPFCMDHICKPVIGTVPVQI---KLAQPGPVLSQAGIPTGSSSKOL 464
Db 692 ----VIOOP-----PKPG-----ALIQLNPLQPPVVPVAVPLP 707
Qy 465 FSLFHVVOOPSGGNEKQVTTTSHSSTLTIOKCGQKTMVNTTIPTSPFPASILKQITLP 524
Db 708 ----VIOOP-----PKPG-----ALIQLNPLQPPVVPVAVPLP 736
Qy 525 GNKIL---SLQASPTQKRIKENVTSCFRDE 552
Db 737 GTKALSAVSAQAAAAQKKNLKEPGGGSFRDD 767
RESULT 14
US-09-724-676A-63922
; Sequence 63922, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; NUMBER OF FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63922
; LENGTH: 1021
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-63922
Query Match 19.1%; Score 527; DB 5; Length 1021;
Best Local Similarity 30.6%; Pred. No. 2.1e-21;
Matches 175; Conservative 58; Mismatches 122; Indels 216; Gaps 24;
Qy 13 PPKVSSG-----PRLPAPQIVAVKAPNTTIOFPANLQLPPTGVLIKSSNGPLM 61
Db 382 PTATTSIGIRATLTPTVLAPRLPQP-----PONPTNIO--NFQLPQGMVLVRSNGQLL 432
Qy 62 LVSPQOQVTR-----AETTSNITSRPVAVPANPQTVKICTVNPSSSOLIKKVAVTPVKKL 115
Db 433 MI-POQALAOQAOAHAPQOTTMARPRATPTSAPVQISTVOAGETPIIAR-QVTP----- 486
Qy 116 AIGTIVTTVPKPSVSVAVPTSVVTPGKPLNTVT--TLKPSISGASS-----TPS 168
Db 487 ----TTIIKQV---SOAQTVPQSAFLOKRSPOVQOLVLGAAQASLGTATAVQGTGPQ 539
Qy 169 NE-PRUKAENSAAVOINISPTMLENVKCKNFAMLIKLAGSGSPKMGONVKKLVEQL 227
Db 540 RTVPGATTSSAATE-----TMENVKCKNFSLTLIKLASSGKOSTETANVKELVONL 593
Qy 228 LDKAKIEEFTFKRLKVELKSSPOPLVPLFKKSVVALRQLLPNSOSFIQOCVOQTSSIMV 287
Db 594 LDGKIEADEFTSKRLRELNSSPOPLVPLFKKSLPALRQLLPDSAAFIQOQQOQPPPP-- 651
Qy 288 IATCTTATTSPPVTTTSSGSEKSIIVSGATAPRTVSVOTLNLAPGVGAKAGAVTLH 347
Db 652 ----TSQATTA--LTAVALSS-----SVQ-----R 670
Qy 348 SVGPAAAGCTTAGGLQTSKPLVTSVANTVTYVLSLOPEKPVVSGTAVTSLPAVTEGE 407

Db 671 TAGKTAAT-----VTSALP--PVLSLT----- 691
Qy 408 TSGAALCLPSVKPVVSPFCMDHICKPVIGTVPVQI---KLAQPGPVLSQAGIPTGSSSKOL 464
Db 692 ----VIOOP-----PKPG-----ALIQLNPLQPPVVPVAVPLP 707
Qy 465 FSLFHVVOOPSGGNEKQVTTTSHSSTLTIOKCGQKTMVNTTIPTSPFPASILKQITLP 524
Db 708 ----VIOOP-----PKPG-----ALIQLNPLQPPVVPVAVPLP 736
Qy 525 GNKIL---SLQASPTQKRIKENVTSCFRDE 552
Db 737 GTKALSAVSAQAAAAQKKNLKEPGGGSFRDD 767
RESULT 15
US-09-724-676-63925
; Sequence 63925, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; NUMBER OF FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63925
; LENGTH: 1053
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-63925
Query Match 19.1%; Score 527; DB 5; Length 1053;
Best Local Similarity 30.6%; Pred. No. 2.2e-21;
Matches 175; Conservative 58; Mismatches 122; Indels 216; Gaps 24;
Qy 13 PPKVSSG-----PRLPAPQIVAVKAPNTTIOFPANLQLPPTGVLIKSSNGPLM 61
Db 414 PTATTSIGIRATLTPTVLAPRLPQP-----PONPTNIO--NFQLPQGMVLVRSNGQLL 464
Qy 62 LVSPQOQVTR-----AETTSNITSRPVAVPANPQTVKICTVNPSSSOLIKKVAVTPVKKL 115
Db 465 MI-POQALAOQAOAHAPQOTTMARPRATPTSAPVQISTVOAGETPIIAR-QVTP----- 518
Qy 116 AIGTIVTTVPKPSVSVAVPTSVVTPGKPLNTVT--TLKPSISGASS-----TPS 168
Db 519 ----TTIIKQV---SOAQTVPQSAFLOKRSPOVQOLVLGAAQASLGTATAVQGTGPQ 571
Qy 169 NE-PRUKAENSAAVOINISPTMLENVKCKNFAMLIKLAGSGSPKMGONVKKLVEQL 227
Db 572 RTVPGATTSSAATE-----TMENVKCKNFSLTLIKLASSGKOSTETANVKELVONL 625
Qy 228 LDKAKIEEFTFKRLKVELKSSPOPLVPLFKKSVVALRQLLPNSOSFIQOCVOQTSSIMV 287
Db 626 LDGKIEADEFTSKRLRELNSSPOPLVPLFKKSLPALRQLLPDSAAFIQOQQOQPPPP-- 683
Qy 288 IATCTTATTSPPVTTTSSGSEKSIIVSGATAPRTVSVOTLNLAPGVGAKAGAVTLH 347
Db 684 ----TSQATTA--LTAVALSS-----SVQ-----R 702
Qy 348 SVGPAAAGCTTAGGLQTSKPLVTSVANTVTYVLSLOPEKPVVSGTAVTSLPAVTEGE 407
Db 703 TAGKTAAT-----VTSALP--PVLSLT----- 723
Qy 408 TSGAALCLPSVKPVVSPFCMDHICKPVIGTVPVQI---KLAQPGPVLSQAGIPTGSSSKOL 464
Db 724 ----VIOOP-----PKPG-----ALIQLNPLQPPVVPVAVPLP 739
Qy 465 FSLFHVVOOPSGGNEKQVTTTSHSSTLTIOKCGQKTMVNTTIPTSPFPASILKQITLP 524
Db 740 ----VIOOP-----PKPG-----ALIQLNPLQPPVVPVAVPLP 768

Thu Feb 20 16:36:34 2003

us-09-763-909-2_copy_1_552.rapn

Page 8

```

Qy      525 GKKIL--SLQASPTQKRNKENVTSCFRDE 552
          | | | | | | | | | | | | | |
Db      769 GPTALSAVSAQAAAAQKKNLKEPGGGSFRDD 799

```

Search completed: February 16, 2003, 22:04:20
Job time : 30.4582 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:55:19 ; Search time 18.6758 Seconds
(Without alignments)
2841.441 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_552

Perfect score: 2758
Sequence: 1 GLLVTKVAPVSAPEKVSQSP.....ASPTQKRIKENVTSCFRDE 552

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	256	9.3	921	2 A48184	transcription init
2	256	9.3	921	2 A45183	TBP-associated fac
3	234.5	8.5	3020	2 A43932	mucin 2 precursor,
4	233.5	8.5	528	2 I47141	gastric mucin (clo
5	226	8.2	1275	2 T33369	hypothetical prote
6	221	8.0	2035	2 A40718	host cell factor C
7	220	8.0	1367	1 S48478	glucan 1,4-alpha-g
8	212.5	7.7	2232	2 T34434	hypothetical prote
9	210	7.6	873	2 A47283	calpoptin - fruit
10	208	7.5	1777	2 T34369	hypothetical prote
11	206.5	7.5	865	2 A47282	calcium-binding pr
12	205	7.4	1032	2 T34433	hypothetical prote
13	204.5	7.4	1324	2 S52863	DNA-binding protei
14	204	7.4	2187	2 T30826	nascent polypeptid
15	201	7.3	873	2 P96615	probable Myb-famil
16	197.5	7.2	5376	2 T42215	zonahesin - mouse
17	197	7.1	648	2 PC4395	mucin 3 - human (f
18	196.5	7.1	3507	2 T45463	membrane glycoprot
19	196	7.1	3507	2 T45413	hypothetical prote
20	195	7.1	1161	2 S57180	probable membrane
21	193.5	7.0	670	2 P36791	hypothetical prote
22	193	7.0	825	2 T29634	hypothetical prote
23	192	6.9	1367	2 S51959	hypothetical prote
24	189.5	6.9	797	1 VGBEX1	glycoprotein x pre
25	188	6.8	866	2 T45462	membrane glycoprot
26	188	6.8	3570	2 T45025	mucin MUC5B, trach
27	186	6.7	1151	2 T18535	high molecular mas
28	185.5	6.7	1630	2 A53577	ascites stialoglyco
29	184.5	6.7	881	2 S56032	probable membrane

30	181.5	6.6	662	2 A45155	mucin FIM-C.1 - Af
31	181.5	6.6	1199	2 A40670	nuclear envelope p
32	181	6.6	1609	2 S25345	probable membrane
33	180.5	6.5	725	2 A41258	a-agglutinin core
34	179.5	6.5	1537	2 S53465	fiocculatin prote
35	176.5	6.4	1260	2 S60896	agglutinin-like pr
36	176.5	6.4	1851	2 T11994	hypothetical prote
37	175.5	6.4	796	2 T21460	hypothetical prote
38	175.5	6.4	2090	2 S26058	hypothetical prote
39	174	6.3	549	2 C87719	probable transform
40	174	6.3	1075	2 S48992	protein R119.6 (im
41	173.5	6.3	583	2 S67571	fiocculatin prote
42	172.5	6.3	851	2 T22696	hypothetical prote
43	172.5	6.3	1858	2 T18273	1-phosphatidylinos
44	172	6.2	886	2 S29605	glycoprotein 350/2
45	172	6.2	1169	2 S38181	fiocculatin prote

ALIGNMENTS

```
RESULT 1
A48184
transcription initiation factor IID 110k chain - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
C:Accession: A48184
R:Kokubo, T.; Gong, D.W.; Roeder, R.G.; Horikoshi, M.; Nakatani, Y.
Proc. Natl. Acad. Sci. U.S.A. 90, 5896-5900, 1993
A:Title: The Drosophila 110-kDa transcription factor TFIID subunit directly interacts
A:Reference number: A48184; MUID:93317591; PMID:8327460
A:Accession: A48184
A:Status: preliminary
A:Molecule type: mRNA; protein
A:Residues: 1-921 <KOK>
A:Cross-References: GB:563550; NID:9398432; PID:9398433
A:Experimental source: embryo nuclear extract
A>Note: sequence extracted from NCHI backbone (NCBIN:134863, NCBIPI:134864)
C:Genetics:
A:Gene: FlyBase:Tafl10
A:Cross-References: FlyBase:FBgn0010280
C:Keywords: transcription Initiation

Query Match          9.3%; Score 256; DB 2; Length 921;
Best Local Similarity 22.9%; Pred. No. 3.8e-07;
Matches 135; Conservative 86; Mismatches 193; Indels 176; Gaps 24;

QY 25 POIYAVKAPNTTIOIPANQLPQTVLIRKNSGFLML-----VSPQQTVAETTS 76
      ||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 133 POSPSTILTSTLNTGOTPA-----LLVKTNDNGHQLLRVGTTGPPVTQTINTSINS 184

QY 77 NITSFANPAPQVYKICTVNSSQ-----LIKVAVTPPKLAQIGTIVTVTP 127
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 185 NITSFTHPTTQ--IRLOTVPAASMTNTTATSNIIIVASVSGVANSQPHLITOLNAO 243

QY 128 KPSSQSVAVPVSVTVPKPLNTVTTLKPSISGASSTPSENPMLKAENSAVQINLSP 187
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 244 APQLPQITQITIPAQSQOOQVNNVSSAGTATVSSTTA-----ATT 287

QY 188 TMLENVK-KCKNFLMLIKLACSGSQSPKMGONKKEVLQDLDAKIEAEFTRLKLYELK 246
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 288 TOOGTKKCKRFLANLIEL--SFREPKVPEKKNVTLLQELVANAVEDEECDRLERLIN 345

QY 247 SSPQHLVPEFLKKSVALRQL-----LPSQSTIQ--- 276
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 346 ASPQCLIGFLKSLPLRLQALYKRELVEIGIKPPQHVGLAGISQQLPIQAOIRDIG 405

QY 277 ---QCVQOTSMDVIATCTTVTTSPPVTTVSSQSEKSIIVSGATAPRTVS---VQT 329
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 406 PSQTTTIGQTVRMI--TPNALGTPRPTIGTTISKQPN---IRLPAPRLVNTGGIRT 460

QY 330 LNPLAGPVGARAVVTLHSVGPRTAATGCTAGTGLQTSKPLVYSVANTVTVTSIQPKP 389
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

Db 461 QIP-SLOVPGQANIYQIR--GPOHQLQRTGSYQIRATRRP-----PNSVPTAN----- 506

QY 390 VVSGTAVTSLPRAVTFGEISGAALICLPBKVAVSFCWDHCKRPVITGTPVQIKLAQEPVL 449

Db 507 -----KLAFAVKGOFOIKAI--TPSLHP-----PSL 530

QY 450 SOPAGIPTGSSSKQLSLFPHVVOQPSGNGEKQVTTISHSSTLTIOKCGQKTMPTVNTIIP 509

Db 531 MAISGPPPTLTSLVS-----TLNSAST-----TTLETPS-LPT 564

QY 510 SOFPASIT---LKQITLPGN-----KLISLQA---SPTQKNRIKENVT 546

Db 565 VHLPPALRAREQMOMSLNHSNHFDAKIVEIKAPSLHPHMERINASTLT 614

RESULT 2

AS183

TFP-associated factor TFIID - fruit fly (Drosophila sp.)

C:Species: Drosophila sp.

C>Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 16-Feb-1997

C:Accession: A45183

R:Hoey, T.; Weidner, R.O.; Gill, G.; Chen, J.L.; Dynlacht, B.D.; Tjian, R. Cell 72, 247-260, 1993

A:Title: Molecular cloning and functional analysis of Drosophila TAF110 reveal properties of Drosophila TFIID subunits

A:Reference number: A45183; MUID:93145326; PMID:7678780

A:Accession: A45183

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-921 <HOE>

A:Experimental source: embryo

A:Note: sequence extracted from NCBI backbone (NCBIP:123832)

C:Genetics:

A:Gene: flybase:Tafl10

A:Cross-references: FlyBase:FBgn0010280

Query Match 9.3%; Score 256; DB 2; Length 921;

Best Local Similarity 22.9%; Pred. No. 3.8e-07;

Matches 135; Conservative 86; Mismatches 193; Indels 176; Gaps 24;

QY 25 POLVAVKANTTTTIOFPANLQDPGTVLTKNSGML-----VSPQVTRAEVTS 76

Db 133 POSPSTLTSLNTGOTPA-----LWKTONGFOLLKRGTTTGPVYQTITNTSNN 184

QY 77 NTSRAVPANPQVTKICVNPSSO-----LKKVAVPVKKLAQIGTTVTVTP 127

Db 185 NNTSTTNHTTTO--IKLQTVPAASMTNTTANSITVNSVASSGVANSSOPHLLTOLMNO 243

QY 128 KSSVQSVAVPTSVTVTPGKPLNTVTLKPSLSGASSTPSNEPNLKAENSAVQINTLSP 187

Db 244 AQLPQITQITPAQOQOQOQVNNVSSAGTATAVSTTA-----ATT 287

QY 188 TMLNENK-KCKNFLMLIKLACSGSSPEMGQNVKVLQDLDAKLEAEFTKLYEIK 246

Db 288 TQOGNFKKCKRFELANLIEL--STREPKVEKVNRTLQELVNAVNEPEFCRRLRLN 345

QY 247 SSPQHLVFLKKSVALROL-----LPNSQSFIO----- 276

Db 346 ASPQCLIFLKKSLPLRLQALYTKELVTEGKPPQHVLAGLSQQLPKIQALRPIG 405

QY 277 ---OCVOQTSMDVATCTTTVTTSPTVTVSSQSEKSIIVSGATAPRVTS---VOT 329

Db 406 PSQTTTIGTQYRMI--TPNALGTPRTTIGHTTISKQPN---IRLPTAPRLVNTGINT 460

QY 330 LNPPLAGPVAKAGVTVLHSGVPTATGTTAGTGLQTSKPLVTSVANTVTVSLOPEK 389

Db 461 QIP-SLOVPGQANIYQIR--GPOHQLQRTGSYQIRATRRP-----PNSVPTAN----- 506

QY 390 VVSGTAVTSLPRAVTFGEISGAALICLPBKVAVSFCWDHCKRPVITGTPVQIKLAQEPVL 449

Db 507 -----KLAFAVKGOFOIKAI--TPSLHP-----PSL 530

QY 450 SOPAGIPTGSSSKQLSLFPHVVOQPSGNGEKQVTTISHSSTLTIOKCGQKTMPTVNTIIP 509

Db 531 MAISGPPPTLTSLVS-----TLNSAST-----TTLETPS-LPT 564

QY 510 SOFPASIT---LKQITLPGN-----KLISLQA---SPTQKNRIKENVT 546

Db 565 VHLPPALRAREQMOMSLNHSNHFDAKIVEIKAPSLHPHMERINASTLT 614

RESULT 3

A43932

mucin 2 precursor, intestinal - human (fragments)

N:Alternate names: mucin SMUC-41

C:Species: Homo sapiens (man)

C>Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999

C:Accession: A49963; A45106; B45106; A43932; B3532; A61257; P00328; P00329

R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S. J. Biol. Chem. 269, 2440-2446, 1994

A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the human MUC2 gene

A:Reference number: A49963; MUID:94132002; PMID:8300571

A:Accession: A49963

A:Molecule type: mRNA

A:Residues: 1-639 <GU1>

A:Cross-references: GB:M94131; NID:9186395; PIDN:AAA59104.1; PID:9186396

R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, J. Biol. Chem. 267, 21375-21383, 1992

A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up and down stream of the mucin domain

A:Reference number: A45106; MUID:93016075; PMID:1400449

A:Accession: A45106

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 626-1895 <GU2>

A:Cross-references: GB:M94131; NID:9186395; PIDN:AAA59104.1; PID:9186396

A:Note: sequence extracted from NCBI backbone (NCBIP:116706)

A:Accession: B45106

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 2037-3020 <GU3>

A:Cross-references: GB:M94131; NID:9186397; PIDN:AAA59104.1; PID:9186398

A:Experimental source: colon

A:Note: sequence extracted from NCBI backbone (NCBIP:116698)

R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, J. Biol. Chem. 267, 21375-21383, 1992

A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorphisms

A:Reference number: A43932; MUID:91558717; PMID:1885763

A:Accession: A43932

A:Molecule type: DNA

A:Residues: 1343-1350, 'L', '1352-1411', 'S', '1413-1448', 'P', '1450-1503', 'T', '1505-1915' <TOR>

A:Cross-references: GB:M74027; NID:9188863; PIDN:AAA59875.1; PID:9188864

A:Note: sequence inconsistent with the nucleotide translation

A:Note: sequence extracted from NCBI backbone (NCBIP:55749; NCBIP:55750)

R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S. J. Biol. Chem. 264, 6480-6487, 1989

A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evolution of the human MUC2 gene

A:Reference number: A43532; MUID:89197956; PMID:2703501

A:Accession: B3532

A:Molecule type: mRNA

A:Residues: 1916-2193 <GU4>

A:Cross-references: GB:M22405; NID:9188873; PIDN:AAA6334.1; PID:9188874

A:Experimental source: Intestine

R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B. J. Clin. Invest. 87, 77-82, 1991

A:Title: Human bronchus and intestine express the same mucin gene.

A:Reference number: A61257; MUID:91086481; PMID:1985113

A:Accession: A61257

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 'T', '1925-1948', 'TTS', '1952-1954' <JAN>

A:Experimental source: bronchus

R:Xu, G.; Huan, L.; Khatri, T.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forslun Blochm. Blophys. Res. Commun. 183, 821-828, 1992

A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the human genome

A:Reference number: P00328; MUID:92198477; PMID:1550588

A:Accession: P00328

A:Molecule type: mRNA

[illegible]

```

Db      852 - -APVSSSTTESSVAVPPTPSSSNITSSAPSIPIFSSTTESFGTGTVPSPSSK ----- 904
OY      255 P L K S V A L R Q L I N S O S F I Q C Y Q O T S D M V I A C T T Y T T S P - - - V T T Y T S S G S E   311
Db      905 ----- Y G S O T - - E T S V S T T E T T I V P K T T I T S V T T P S T T I T T Y T C S G T G N   949
OY      312 K S I I V S G A T A P R T V - - S V Q I L N P L A G P V A K A G V T T L H S V P P T A T G T T A G T G - - L I Q T   367
Db      950 S A G E T T S G C S P R T V T T Y P T T T T S V T T S T T T I T T Y C S G T N S A G E T T S G C S P K I T T Y   1009
OY      368 S K P L A T S V A N T - - - V T T V S I Q E K P V S G R A V T L S L P A V T - - F G E T S G A A I C - - - - -   414
Db      1010 T V P C T S E S E T S E S T T S P T T P V T T V T T V T T E Y S T K P G E I T T T E V T K N I P T Y   1069
OY      415 ----- L S Y K P V S F C M D H I C K P Y I G T P V O I K A Q P P V L S Q A G I P T G S S Q L F S L F   468
Db      1070 L T T I A P T S P S V T T V N F ----- P P T I T T T - - - V C S - - - - T G N S - - - - -   1101
OY      469 H V W Q P S G G N E K O V T T I S H S T L T I Q C G O K T M P V N T I I P S   510
Db      1102 - A G E T T S G C S P R T Y T T Y P C S T G T - - - G E Y T T E A T I L Y T T A   1138

RESULT 8
t34434
hypothetical protein K06A9_1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: t34434
R:Geisel, C.; Gatlung, S.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: Z21525
A:Accession: t34434
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2232 <GEI>
A:Cross-references: EMBL:U08046; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9_1a
A:Experimental source: strain Bristol NZ; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9_1a
A:Map position: X
A:introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 20

Query Match          7.7%; Score 212.5; DB 2; Length 2232;
Best Local Similarity 21.9%; Pred. No. 0.00039;
Matches 138; Conservative 73; Mismatches 247; Indels 173; Gaps 24;

OY      26 Q I V A V K A P N T T I O P A M I Q L P P G V L I K S N G P M L V S P Q O T V R A E T T S N I T S R P A V P   85
Db      801 Q S V T S N S F G S T Y T R ----- P S T V S G T S G S T V Y V G S T E A S T S G S S V A S S P A P S T S   852
OY      86 A N P O T V K I C T V P N S S S O L I K K V A V T P V K K L O I G T V V T T V P K P S V Q S V A V P S V V T V T   145
Db      853 Q N P N ----- P S T S ----- G S M I T G S P Y S - Q S T S P V A S S T T P S   886
OY      146 P E K P L A N T Y T T L K P S ----- S L A S S T P S N ----- E P N L   173
Db      887 P E S P E T T I T S P S P S O S T T I G S O G S T P G I S T T S E K M S O G S T O P G S T G S T V T O P S T   946
OY      174 K A E N A A - - V O I N S ----- P T M L E N V K K C N F L A M L I K L A C S G S - - - G S P E M G N   219
Db      947 V S D S T S S S T Y T V G S T E B S S S P I P S T S O N T P S T S - - - - - S G S M S T O T P O S S O S   996
OY      220 V K I I E O L L D A K I E A E F T R K L Y E L K S S P O P H - - - - - P N - S O S F I Q C Y Q O T S S D M V I A   289
Db      997 T S P - V E S T S G A T S S G S G P T T L N I S P S P E S S T I G S S O G S T S P V V S T I S O G S T E P G S   1055
OY      258 K K S V A L A Q L L - - - - - P N - S O S F I Q C Y Q O T S S D M V I A   289
Db      1056 T G S Y T T K P S T V G S A S G S G T A T M G S T E A S T S G S S T S P N S O S T S P S T S G A T S S P S S G   1115

```


Db 926 TSAPQSSSTSAATVYSSS-----QSPTSSPAQSSSTPAQSSSTVVOSSSFQ 975
Qy 498 --QKTMV--NTIIPQSOPPASI-----LKOITLPG--NKILSLQASPTOKMRI 541
Db 976 SPOSTIGSSTVPSIOPASSSTSGPTTQICPNQOTVFKGQVGIYEMLPASTQONAI 1035

Qy 542 K---ENV 545
Db 1036 NAFVENY 1042

RESULT 11

A47282
calcium-binding protein calphotin - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C:Accession: A47282
R:Martin, J.H.; Benzer, S.; Rudnicka, M.; Miller, C.A.
Proc. Natl. Acad. Sci. U.S.A. 90, 1531-1535, 1993
A:Title: Calphotin: a Drosophila photoreceptor cell calcium-binding protein.
A:Reference number: A47282; MUID:93165729; PMID:8094559
A:Accession: A47282
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-865 <MAR>
A:Cross-references: GB:L02111; NID:g157031; PIDN:AAA28405.1; PID:g157032
A:Experimental source: photoreceptor cells
A:Note: sequence extracted from NCBI backbone (NCBIN:124955, NCBIPI:124956)
C:Genetics:
A:Gene: FlyBase:Cpn
A:Cross-references: FlyBase:FBgn010218
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: calcium binding

Query Match 7.5%; Score 206.5; DB 2; Length 865;
Best Local Similarity 21.7%; Pred. No. 0.00025;
Matches 122; Conservative 82; Mismatches 242; Indels 115; Gaps 22;

Qy 1 GTLVTKV-APVSAP-----PKVSSGRLPAP---QIVAVKAPNTTIOFPANLQ 45
Db 4 GTISPSAPVAPVYTSAAVAPVQVSPAAVAPAPVAPVAPVAPVAPVAPVAPVAPV 63
Qy 46 LPPTVLIKNSGPMLVSPQOTVTRAETTSNTRSPAVPAPQTVKICVTPNS-SSQLI 104
Db 64 IPAPAPIAASVTPVAVSAP--PVVAPTP-----PAASVSTPVAVAQIPVAVSAPVA 115
Qy 105 KKVAVTTPVKL-AGITVTVTVKPSVSVAVPTSVT-VTGRKPLNTYTTLKPSLQ 162
Db 116 PPAVATPTPVQVLPVAPVATTPVAAAPPAVATPVISPVVAPVAPVAPVAPVAPV 175
Qy 163 ASSTPSNEPILKENSAAVOINLSPMLE-----NYKCKKNFLMLIKLACS 209
Db 176 VAAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAV 227
Qy 210 GSGSPENGQNVKLVLEQLDAKEEFTKRLYELKSSPQHLVPLFKSVVALROLPL 269
Db 228 ----PEVSVATKPLAAEPVAVPATEPVPVAPAAASPHVAVAPAVETVAVA-----P 278
Qy 270 NSQSFIOCCVQOQSSDMVATCTTTVTSPVYTTVSSSQSEKSIYSGATAPRTVSQY 329
Db 279 VVAS-----TEPPVAAATLTAPET-PALAVVAVESQVAVANTVAVATPPTP-APPEET 328
Qy 330 LNPAGPVGAKAGVVLTHSVGPTAAGTGTAGTGLDTSKPLSVANTVTVTSLOPEKP 389
Db 329 IAP-----PVAETPEVAVSAVAETTPVVPVA-----AESIPAP 364
Qy 390 VSGTAATVLSLPAVTFETSGAALCLPVRKVS-----FCMDHICKPVYGT 436
Db 365 VAVATPVPAVL-AVTDVDAASAV--PELPPVIAVSPVPAVPAVPAVPAVPAVPAV 421
Qy 437 PVQIKLAPGVPVLSQPGVIGIPGSSSKQLFLSHVAVQDPSGGNEKQVTTISHSSTLTQKC 496

Db 422 PVPAVVAEETPEPAPASAPVTLAALDIEVAVPAVPAVPAVPAVPAVPAVPAV 473
Qy 497 GQKTMVNTIIPQSOPPASI 517
Db 474 -VSTPPTTASVPETTAPPAV 493

RESULT 12

T34433
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34433
R:Giesel, C.; Gattung, S.
submitted to the EMBL data library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: Z21525
A:Accession: T34433
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1032 <GEI>
A:Cross-references: EMBL:U00846; PIDN:AC70889.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 839/1; 849/1; 875/3; 940/2; 966/2;
Query Match 7.4%; Score 205; DB 2; Length 1032;
Best Local Similarity 22.7%; Pred. No. 0.00038;
Matches 133; Conservative 74; Mismatches 227; Indels 152; Gaps 24;

Qy 2 TLVAVPAVSPRVSSGRLPAP---QIVAVKAPNTTIOFPANLQPLPCTVLIKNSG 58
Db 246 SITSALPT-ASSASSPASASTTPVLLSSTIOSSSGTFPSSVASSPSTV--GSTSG 302
Qy 59 PLMLVSPQOTVTRAETT-SNTRSPAVPAPQTVKICVTPSSQLIKKAVTPEKKLQ 117
Db 303 AASSSVATVSTIAGSGSTIT---PVGSSSTIGSSSTPMASSSSSGTMSIS-----GS 354
Qy 118 IGTIVTVTVKPS-----VQSVAVPTSVYVTPGKPLNTVTTTLKPSLIGASPTBNE 170
Db 355 TGST-VTVVPGSSSTFASSTPIASSSGSTVTAFGS-----SSTYSGSTPS-- 401
Qy 171 PNLKAENSAVQINLSPMLENVKCKNFLMLIKLACSGSGSPDMGQNVKLVLEQLDA 230
Db 402 -----ASSSSSGTMSNGSTGTYVA 424
Qy 231 KIEAEFTKRLYELKSSPOHLPFLKSVVALROLPLNSQSFIOCCVQOQSSDMVAT 290
Db 425 PVSSSTFG-----SNTP-----IASSSSSGSTVTVVSGSSSTYGSS 460
Qy 291 CTTVTTSPTVYTTVSSSQSEKSIYSGATAPRTVSQTLNP-----LAGVGAAG 342
Db 461 TPSASSSAGTASTISGTSATIVPSSSSVSSSTQASAPSSPGTMTSTVSGPTGTVT 520
Qy 343 VTLHSVGP-----AATGTTAGTGLDTSKPLSVANTVTVTSLOPEKPV 391
Db 521 VVPGSSSTPAPSSSPNPSSSPASTGSTITITISG-----SSIIIVTIVSGS-----TV 566
Qy 392 SGAATVLSLPAVTFETSGAALCLP-SVKPVVSCMDHICKPVYGT--TPVQIKLAPGPV 448
Db 567 SGTSTQSGSTLASSTATPGSSSTVPSSSSPQS---SQSPAPNMGSTTPGOTSSQSPSPS 623
Qy 449 LSGAPGIPGSSSKQLFLSHVAVQDPSGGNEKQVTTISHSSTLTIT--OKCGQKTMVNTI 506
Db 624 MNPSSSTPTGSSGSTITPEGSTASSPTG-----STGSTFVAVTEVTSQSTVPSSGSS 674
Qy 507 I---PTSGPPASILKQITLPGKNKILSLQASPT--TOKNRKIENTVS 547
Db 675 LGTQSTNSPSPSSLSPT-SGMSTLTSEPSSTQSGAOSTLTIT 719

...

...

...

...

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:54:50 ; Search time 10.6381 Seconds

(without alignments)
2152.161 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_552

Perfect score: 2758

Sequence: 1 GTLVTKVAPVAPKVSQSGP.....ASPTOKRIKENVTSQFRDE 552

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match Length	ID	Description
1	2758	100.0	801	T2DT_HUMAN	Q92750 homo sapien
2	547	19.8	1083	T2D3_HUMAN	Q00268 homo sapien
3	256	9.3	921	T2D3_DROME	P47825 drosophila
4	223.5	8.1	2035	HFC1_HUMAN	P51610 homo sapien
5	223	8.1	2090	HFC1_MESAU	P51611 mesocricetu
6	220	8.0	1367	AME1_YEAST	P08640 saccharomyc
7	219	7.9	5179	MUC2_HUMAN	Q02817 homo sapien
8	206.5	7.5	865	CPN_DROME	Q02910 drosophila
9	197.5	7.2	5376	ZAN_MOUSE	O88799 mus musculu
10	195	7.1	1161	DAN4_YEAST	P47179 saccharomyc
11	195	7.1	1322	YAG3_YEAST	P39712 saccharomyc
12	194.5	7.1	2700	ZAN_HUMAN	O9Y493 homo sapien
13	193.5	7.0	670	VG50_HSV1	Q00130 ictaluriid h
14	189.5	6.9	797	VG1X_HSV1	Q00130 ictaluriid h
15	184.5	6.7	881	JUH8_YEAST	P28968 equine herp
16	181.5	6.6	662	MUC1_XENLA	P47033 saccharomyc
17	181.5	6.6	1199	N121_RAT	O05049 xenopus lae
18	181	6.6	1609	FTG2_YEAST	P25591 rattus norv
19	180.5	6.5	725	AGAL_YEAST	P32323 saccharomyc
20	180.5	6.5	3178	YS89_CABEL	Q09624 caenorhabdi
21	179.5	6.5	1537	FL01_YEAST	P32768 saccharomyc
22	178.5	6.5	1260	AL51_CANAL	P46590 candida alb
23	175.5	6.4	2090	N214_HUMAN	P35658 homo sapien
24	174	6.3	1075	FLOS_YEAST	P38894 saccharomyc
25	172.5	6.3	1858	P3K2_DICDI	P54674 dictyostell
26	172	6.2	886	VG3_EBYA8	O07284 epstein-bar
27	172	6.2	1169	YK82_YEAST	P36170 saccharomyc
28	169.5	6.1	745	OCT1_PIG	Q29076 sus scrofa
29	169	6.1	1229	N121_HUMAN	O9Y263 homo sapien
30	168	6.1	1306	MSB2_YEAST	P32334 saccharomyc
31	167.5	6.1	1140	YMG6_YEAST	O04883 saccharomyc
32	166.5	6.0	743	OCT1_HUMAN	P14859 homo sapien
33	166.5	6.0	3866	HRX_MOUSE	P55200 mus musculu

34	166	6.0	3726	1	TRX_DROME	P20659 drosophila
35	165	6.0	606	1	SP2_HUMAN	Q02086 homo sapien
36	164.5	6.0	1365	1	SUZ2_DROME	P25172 drosophila
37	163	5.9	860	1	CH12_COCIM	P54197 coccidioides
38	162	5.9	3969	1	HRX_HUMAN	O03164 homo sapien
39	161	5.8	1119	1	AL53_CANAL	O74623 candida alb
40	161	5.8	1509	1	GSRI_HUMAN	O9ZM04 homo sapien
41	160.5	5.8	630	1	MUC1_MOUSE	O02496 mus musculu
42	159	5.8	676	1	MUC1_MESAU	O60528 mesocricetu
43	157.5	5.7	907	1	VG3_EBY	P03200 epstein-bar
44	153	5.5	739	1	OCT1_CHICK	P15143 gallus gall
45	152	5.5	604	1	MTG8_HUMAN	O06455 homo sapien

ALIGNMENTS

```

RESULT 1
ID T2DT_HUMAN STANDARD; PRT; 801 AA.
AC Q92750:
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE Transcription initiation factor TFIID 105 kDa subunit (TAFII-105)
DE (TAFII105) (Fragment).
GN TAF4B OR TAF2C2 OR TAFII105.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=97011146; PubMed=8858156;
RA Dikstein R., Zhou S., Tjian R.;
RT "Human TAFII 105 is a cell type-specific TFIID subunit related to
   hTAFII130."
RL Cell 87:137-146(1996).
CC -1- FUNCTION: CELL TYPE-SPECIFIC SUBUNIT OF TFIID THAT MAY FUNCTION AS
   A GENE-SELECTIVE COACTIVATOR IN CERTAIN CELLS. TFIID IS A
   MULTIMERIC PROTEIN COMPLEX THAT PLAYS A CENTRAL ROLE IN MEDIATING
   PROMOTER RESPONSES TO VARIOUS ACTIVATORS AND REPRESSORS.
CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
   NUMBER OF TBP-ASSOCIATED FACTORS (TAFs). TAFII105 MAY FORM
   HETERODIMERS WITH TAFII130.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: FOUND IN B CELLS BUT NOT IN OTHER TESTED
   CELLS.
CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC EMBL: Y09321; CA70499.1; -
CC Genew; HGNC:11538; TAF4B.
CC MIM; 601689; -
CC InterPro: IPR003894; TAF_hom.
CC SMART: SM00549; TAFH; 1.
CC Transcription regulation; Nuclear protein.
CC NON_TER 1
SQ SEQUENCE 801 AA; 85658 MW; D1284932FEA9CD2 CRC64;

Query Match 100.0%; Score 2758; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 8.3e-147;
Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 GTLVTKVAPVAPKVSQSGRLPAPOIVAVKAPMTTITQFPANILQPLPGTGLVLIKNSGSL 60
|||||

```

```

Db 1 GRLVTKVAVSAPKRVSSGRLPAPQIVAVKAPNTTIOFANLQLPSTGLIKNSGRL 60
Oy 61 MLYVPOQVYTRARETTSNTSRAPVAPNPQVATCTVPSNSQOLIKKVAATPVAKKLAQIGT 120
Db 61 MLYVPOQVYTRARETTSNTSRAPVAPNPQVATCTVPSNSQOLIKKVAATPVAKKLAQIGT 120
Oy 121 TVTAVTPKRSVQSVAVPTSVTPGKPLNTVTTLKPSLSGASSTPSNEPNLKANSNA 180
Db 121 TVTAVTPKRSVQSVAVPTSVTPGKPLNTVTTLKPSLSGASSTPSNEPNLKANSNA 180
Oy 181 VOINLSPTMLENVKKCKNEFLAMLIKACSGSSPEMGQNVKRLVQLDLDAKTEAEETFRK 240
Db 181 VOINLSPTMLENVKKCKNEFLAMLIKACSGSSPEMGQNVKRLVQLDLDAKTEAEETFRK 240
Oy 241 LVELKSSQPHLPVPLFKSSVVALQQLPNSQSFIOQCVOQSSSDPVATCTTTVTTSV 300
Db 241 LVELKSSQPHLPVPLFKSSVVALQQLPNSQSFIOQCVOQSSSDPVATCTTTVTTSV 300
Oy 301 VTTVSSSQSEKSIIVSGATAPRTVSVOTLNPVAGVAGVATLHSGPTAATGTTA 360
Db 301 VTTVSSSQSEKSIIVSGATAPRTVSVOTLNPVAGVAGVATLHSGPTAATGTTA 360
Oy 361 GTGLIQTSPKPLVSVANTVTVSLQPEKPVSGTAVTSLPAVTEGTSGAALICLPSVXP 420
Db 361 GTGLIQTSPKPLVSVANTVTVSLQPEKPVSGTAVTSLPAVTEGTSGAALICLPSVXP 420
Oy 421 VVSFPCMDHCKRPGVIGTPVQIKLAPGCPVLSOPAGIPTGSSSQQLPSLFHVQOOPSGGNK 480
Db 421 VVSFPCMDHCKRPGVIGTPVQIKLAPGCPVLSOPAGIPTGSSSQQLPSLFHVQOOPSGGNK 480
Oy 481 QVTTTSHSSTLTIOKCGQKTPMPTIIPGPPASILKQITLPGKTILOASPTQKR 540
Db 481 QVTTTSHSSTLTIOKCGQKTPMPTIIPGPPASILKQITLPGKTILOASPTQKR 540
Oy 541 IKENVTSCFRDE 552
Db 541 IKENVTSCFRDE 552

```

RESULT 2
ID T2D3_HUMAN STANDARD: PRT: 1083 AA.
AC 000268; Q99721; Q9BR42; Q9BR40;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription Initiation factor TFIIID 135 kDa subunit (TAFII-135)
DE (TAFII135) (TAFII-130) (TAFII130).
GN TAF4 OR TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=97336072; PubMed=9192867;
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blake S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Buttrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Cordy N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,

```

RA Graffam D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.D.,  
RA Humble E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
RA Levasaio M.B., Leverha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
RA Marsh V.L., Martin S.L., McConachie L.J., McIay K., McMurray A.A.,  
RA Milne S.A., Misty D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
RA Phillimore B.J.C.T., Prichaltingam S.R., Plumb R.W., Ramsay H.,  
RA Rice C.M., Ross M.T., Scott C.E., Selha H.K., Showkeen R., Sims S.,  
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
RA Tracey A., Tromans A.C., Vaudin M., Walli M., Wallis J.M.,  
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,  
RA Wilting L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
RA Rogers J.;  
RT "The DNA sequence and comparative analysis of human chromosome 20.";  
RL Nature 414:865-871(2001).  
RN (3)  
RP SEQUENCE OF 105-1083 FROM N.A. AND PARTIAL SEQUENCE.  
RX MEDLINE=97098442; PubMed=8942982;  
RA Tanese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;  
RT "Molecular cloning and analysis of two subunits of the human TFIIID  
complex: hTAFII130 and hTAFII100.";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).  
CC - FUNCTION: MAKES PART OF TFIIID IS A MULTIMERIC PROTEIN COMPLEX THAT  
CC PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS  
CC ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION  
CC BY THE AP-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.  
CC - SUBUNIT: TFIIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A  
CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).  
CC - SUBCELLULAR LOCATION: Nuclear.  
CC - SIMILARITY: BELONGS TO THE TAF2C FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL: Y11354; CAA72189.1; -  
DR EMBL: AL137077; CAC36006.1; -  
DR EMBL: AL109911; CAC22312.2; -  
DR EMBL: U75308; AAC50901.1; -  
DR TRANSFAC: T02328; -  
DR GeneW: HGNC:11537; TAF4.  
DR MIM: 601796; -  
DR InterPro: IPR003894; TAF_hom.  
DR SMART: SM00549; TAFH; 1.  
KW Transcription regulation; Nuclear protein.  
KW DOMAIN 39 42  
FT DOMAIN 52 57  
FT DOMAIN 98 101  
FT DOMAIN 142 148  
FT DOMAIN 268 275  
FT DOMAIN 331 337  
FT DOMAIN 680 683  
FT DOMAIN 808 813  
FT DOMAIN 828 831  
FT DOMAIN 105 117  
FT CONFLICT 136 136  
FT CONFLICT 185 185  
FT CONFLICT 233 264  
FT CONFLICT 293 293  
SQ SEQUENCE 1083 AA; 109943 MW; A645382575A0752 CRC64;  
Query Match 19.8%; Score 547; DB 1; Length 1083;  
Best Local Similarity 30.2%; Pred. No. 2,4e-23;  
Matches 173; Conservative 67; Mismatches 144; Indels 188; Gaps 21;

```


Query Match 9.3%; Score 256; DB 1; Length 921;
 Best Local Similarity 22.9%; Pred. No. 3.3e-07;
 Matches 135; Conservative 86; Mismatches 193; Indels 176; Gaps 24;

25 PQIVAVKAPNTTITQPANLQPPGVILKSNSSPLML-----VSPQOVTTRAEVTS 76
 133 PPSSTITSLTNGQPR-----LVKTDNGFOLLRGVTTGPTVQTITNTSNN 184
 77 NITSRVAVANPQTVKICVYPNSSQ-----LTKKAVIPVKLAQIGTVVTVTP 127
 185 NITSTNNHPPTTQ-IRLQVYVPAASMTNTTATNTIIVNSVASSGVANSSQPHLTOLMAQ 243
 128 KPSSVAVAPTSVTVYTPGKPLNTVTTKPSLSGASPSFSPNPKLKAENSAVQVNLSP 187
 244 ARLPQITQITQIPAAQSOQOQVNVNSAGGTAVAVSSTA-----ATT 287
 188 TMLENVK-KCKNFLMLIKLACSGSSPENGQVKKLVEOLDLAKTEAEERKLYVELK 246
 288 TQGNTRKCKRKLAMLIEL--STRBPKPYEKVNRLLQELVANAVEPEEFCRLERL 345
 247 SSFQPHLVFFPKKSVVALQL-----LPNSQSFIO--- 276
 346 ASPOQLIFLKKSLPLRLQALYKELVLEGIRPPQHVLAGLSQQLFKIAQAIRPIG 405
 277 ---QCVQSTSDMTATCTTIVTSPVYTTVSSQSEKSIIVSGATAPRTVS---VOT 329
 406 PGGTTTIGQTVAMT--TPRALGTPRTTIGHTTISKOPR---IRLPTRALVNTGIGRT 460
 330 LNPAGPVAKAGVNTLHVSVPAAATGCTAGTGLQTSKPLVTSVANTVTSLOPEK 389
 461 QLP-SLVQFGQANIVQIR--GPHQALQRTGSGVQIRATRP-----PNSVPTAN----- 506
 390 VVSGTAVTSLPVTGEGSGAICLPVSKPVVSPFMDHICRPVIGTPVQIKIAQEPVL 449
 507 -----KLTAVKVGQIQIKAI-TPSLHP-----PSL 530
 450 SOPAGIPTSSSKQSLFLFHVYQPSGNEKQVTTISHSSTLTQKCGKTMPTVNTIIP 509
 531 AATSGPPTPTLSTVLS-----TLNSAST-----TTLPDIPS-LPT 564
 510 SOPPAST---LKQITLPGN-----KILSLQA---SPTOKRIKENVT 546
 565 VHLPEALRAREQMONSLNHSNHFDKIVEIKAPSLHPHMERINASTL 614

RESULT 4
 ID_HUMAN STANDARD; PRT; 2035 AA.
 AC P51610;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Host cell factor C1 (HCF) (VP16 accessory protein) (HFC1) (VCAF)
 GN HCF1 OR HCF1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS), AND PARTIAL SEQUENCE.
 RX MEDLINE=93327419; PubMed=8392914;
 RA Wilson A.C., Lamarco K., Peterson M.G., Herr W.;
 RT "The VP16 accessory protein HCF is a family of polypeptides processed
 from a large precursor protein.";
 RL Cell 74:115-125(1993).
 RC [2]
 RP SEQUENCE OF 65-2035 FROM N.A. (LONG AND SHORT ISOFORMS).
 RC TISSUE=Fetal brain;
 RA MEDLINE=95130107; PubMed=7829097;
 RA Patrosso M.C., Strina D., Susani L., Vezconi P.;
 RT "Genomic organization of the human VP16 accessory protein, a

housekeeping gene (HCF1) mapping to Xq28.";
 RT Genomics 23:30-35(1994).
 RN [3]
 RP PROCESSING.
 RX MEDLINE=96033796; PubMed=7590226;
 RA Wilson A.C., Peterson M.G., Herr W.;
 RT "The HCF repeat is an unusual proteolytic cleavage signal.";
 RL Gene Dev. 9:2445-2458(1995).
 CC -1- FUNCTION: UPON LYTIC INFECTION OF PERMISSIVE CELLS, THE HSV
 CC TRANSCRIPTIONAL PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF
 CC ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING
 CC PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR
 CC ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES.
 CC -1- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM
 CC 110 TO 150 KDa AND A MINOR 300 KDa POLYPEPTIDE. THE MAJORITY OF N-
 CC AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBERT
 CC NONCOVALENTLY, ASSOCIATED.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 CC SHORTER FORM, MAY BE PRODUCED BY ALTERNATIVE SPLICING. IN THE
 CC SHORTER FORM THE N- AND THE C-TERMINAL FRAGMENTS FAIL TO
 CC ASSOCIATE.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN FETAL TISSUES AND THE
 CC ADULT KIDNEY. PRESENT IN ALL TISSUES TESTED.
 CC -1- DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE
 CC SIGNAL.
 CC -1- PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH
 CC CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS
 CC AT A DEFINED SITE, PCE/HER, WITHIN THE HCF REPEAT.
 CC -1- PTM: GLYCOSYLATED. CONTAINS O-LINKED N-ACETYLGLUCOSAMINE RESIDUES.
 CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC -----
 DR EMBL: L20010; -; NOT_ANNOTATED_CDS.
 DR EMBL: X79198; CAA55790.1; ALT_INIT.
 DR Gene: HGNC:4839; HCF1.
 DR MIM: 300019; -
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR001798; Kelch.
 DR Pfam: PF00041; fn3; 1.
 DR Pfam: PF01344; Kelch; 5.
 DR SMART: SM00060; FN3; 1.
 KW Nuclear protein; Repeat; Alternative splicing; Glycoprotein.
 FT REPEAT 44 89
 FT REPEAT 93 140
 FT REPEAT 148 194
 FT REPEAT 217 265
 FT REPEAT 266 313
 FT DOMAIN 1010 1439
 FT REPEAT 1010 1035
 FT REPEAT 1072 1097
 FT REPEAT 1101 1126
 FT REPEAT 1158 1183
 FT REPEAT 1286 1311
 FT REPEAT 1314 1339
 FT REPEAT 1349 1374
 FT REPEAT 1414 1439
 FT REPEAT 382 450
 FT VARSPLIC 564 564
 FT CONFLICT 603 603
 FT CONFLICT 1164 1164
 FT CONFLICT 1873 1873
 SQ SEQUENCE 2035 AA; 208841 MW; 99207FBB875204C0 CRC64;

Query Match 8.1%; Score 223.5; DB 1; Length 2035;
 Best Local Similarity 22.9%; Pred. No. 5.7e-05;

Matches 147; Conservative 79; Mismatches 217; Indels 199; Gaps 33;

QY 2 TLVTKAVAPSPKRVSSGRPLP-----APQIVAKAPNTT-----IQ 39

Db 446 TLVTPQAP--APPTTTTIOVLPYRGSSISVPATKQSPAVAKVGPATTTTPIVLM 503

QY 40 FPN-----LQDPGVL-----KNSGPMVSPDOT-----VTRAETTSNI--TS 80

Db 504 RPASQAKAVVTVSLPAGVAVVPTQSAQGVTVIGSSPOMSGMAALAAAAATOKIPSS 563

QY 81 RP--AVPN-----PQVAKICVP-----NSSQILKKVAVTVKLAQ 117

Db 564 RPYVLSVPACTITVKMAVTPGTTLPATVAVKSSPMVSPATRMKTA-----AQ 616

QY 118 IGTTV-----VTVKRPSSQSAVAVTVTPGKPLMTVTLK--PSSL--GASS 165

Db 617 VGRSVSATVMTSTRTPIITVHKSGTV--TVAGQAVTVTVGVTMTITLVKSPISVPGSA 675

QY 166 TPNSEPL-----KAENSAVQINLSPTMLENKKCKNPL--AMLIKLACSGSQSP-- 214

Db 676 LISMLKVMVSVQTKPVQTSVAVTGOASTGCVTOIITQKGPLPAGTITIKLVTSADCKPPTI 735

QY 215 -----EMGQVVKRLVEQLLAKIEEFTRKLYELKSPQPHVAPPLKKSVALRQLP 269

Db 736 ITTQASGAGTKPIIGI-----SSVSPSTT--KPGTTTITITP 773

QY 270 NSQSFIOQCVQOTSSDVAVIATCTTVT-----SPVVT-----TVSSSQSEKSI 314

Db 774 MSALITQAGATGVYSSSGIKSPIITITTKVMTSGTGAAPAKIITVAPKATIGHQGGQVQV 833

QY 315 IVGA-----TAPRT-----VSOTLNLAPGVAK--AGVYTLHSVGPATAG 356

Db 834 VLKAGAPQCPITLRTVPMGCVRLVTPVAVKPAVTVLVKGTGTGTLCTGVTVST- 892

QY 357 GTTAGLLQTSKPLVSV--NTVTVSLQPEKPVYSGRAVTLSTAVPFGESGAIC 414

Db 893 -SLAGAGHSTSLAPITTLGITATLSSO--VINPAITVSAAGTTLTAAGLTPP 947

QY 415 LPSKVPVSPFCMDHICKPVIGTVPQIKL--AOPGVLSQPA--GIPTGSSSKQLFSLFHVQ 472

Db 948 TITMGPV-----SOPVQVTLITAPSGVEAQPVDHLPV-----ILA 983

QY 473 QPSGGENKQVTTISHSTLTIOCKGQKTPMVTIIPTSQPPP 514

Db 984 SP-----TTQPTATVITADSGQGVQPGTTLVCSNP 1017

RESULT 5

HFC1_MESAU STANDARD; PRT; 2090 AA.

ID HFC1_MESAU STANDARD; PRT; 2090 AA.

AC P51611;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Host cell factor C1 (HCF) (VP16 accessory protein) (HFC1) (VCAF) (CEF).

GN HCF1.

OS Mesocricetus auratus (Golden hamster).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Mesocricetus.

OC NCB1_TaxID=10036;

OX NCB1_TaxID=10036;

RN [1]

RP SEQUENCE FROM N.A.

RA Goto H., Nishitani H., Umene K.I., Nakabeppu Y., Nishimoto T.;

RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: UPON LYTIC INFECTION OF PERMISSIVE CELLS, THE HSV TRANSCRIPTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES (BY SIMILARITY).

CC -1- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM

CC 110 TO 150 kDa AND A MINOR 300 kDa POLYPEPTIDE. THE MAJORITY OF N- AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBEIT NONCOVALENTLY, ASSOCIATED (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE SIGNAL (BY SIMILARITY).

CC -1- PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS AT A DEFINED SITE, PCEC/THER, WITHIN THE HCF REPEAT (BY SIMILARITY).

CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; D45419; BAA08258.1; -.

DR Interpro: IPR003961; FN1.III.

DR Interpro: IPR001798; Kelch.

DR Pfam; PF00041; Fn3; 2.

DR Pfam; PF01344; Kelch; 5.

DR SMART; SM00060; FN3; 2.

KW Nuclear protein; Repeat.

FT REPEAT 44 89

FT REPEAT 93 140

FT REPEAT 148 194

FT REPEAT 217 265

FT REPEAT 266 313

FT DOMAIN 1010 1448

FT REPEAT 1010 1035

FT REPEAT 1072 1097

FT REPEAT 1101 1126

FT REPEAT 1157 1182

FT REPEAT 1295 1320

FT REPEAT 1323 1348

FT REPEAT 1358 1383

FT REPEAT 1423 1448

FT SEQUENCE 2090 AA; 214942 MW; E495EB81F385E17 CRC64;

Query Match 8.1%; Score 223; DB 1; Length 2090;

Best Local Similarity 22.0%; Pred. No. 6,2e-05;

Matches 130; Conservative 75; Mismatches 190; Indels 196; Gaps 27;

QY 10 VSAPKVVSSGRLPAPQIVAKAPNTTIOFPANLQLEPGTVLIKNSGPMVSPQOTV 69

Db 537 IGSNPOKSGMAALAAAAATOKIPSSA--PTVLSVPAAGTIVKT-----VAVTPTT 588

QY 70 TRAFETSIITSRPAVPANPQ-----VAICVPPNSSQ--LIKVAATPVKKL 115

Db 589 LPA--TVKAVASSPVAVSPATRMKLTAAAOVGTSSSAANTSTRTIIVHSGTVYAAQ 646

QY 116 AQISTV-----TVPKP-----SSVQVAVPTS-----140

Db 647 AQVTVVGGVTKRITLVKSPISVPGSALISNLKVMVSVQTPPVQTSVAGASTGPV 706

QY 141 -----VTVTPGKPLMTVTL-----KPSLSGAST-----PSNE 170

Db 707 TQIOTGKPLPAGTILKLVTSADGKPTIITTTQASGASKPITLIGISSVSPSTKPGTT 766

QY 171 PNLAENSAVQINLSPTMLENKKCKNPLML--IKLACSSQSPKQNVKRLVEDL 229

Db 767 TITTPMSALITQAGATGVSTPGIKSPIITITTKVTSCTGAP-----AKITAV-- 818

QY 230 AKIAEEFTRKL--VYELKSSP--OPHLVFLKKSVALROLPPNSQSFIOQCVQOTSSDMV 287

Db 819 PKATGQCGCVTVVLKAGAGQG-----ALIKTV-----MSGVR 855

QY 288 IATCTTIVTTSVTVTVSSQSEKSIIVSGATAPRTVSVQTLNPLAGPVGAKAGVYTLH 347

```

Db 856 LTPVTVSAVKRAVTT-----LVYKGT-----GVTTLG 884
Qy 348 SVCPPTAANGCTTACGGLQTSKPLVTSVA--NTVTVSLQPEKPVVSGTAVTLSPAVTF 405
Db 885 YVTGVST--SLAGAGAHSTASLAPITTLGTLATLSL-----VINPATAVSAOTLL 938
Qy 406 GETSGAICLPKPVVSCWCHICKPVIGTPOIKL-AQPGVYLSQPA-GIPTGSSSKO 463
Db 939 TAAGGTLTPPTTMOV-----SOPQVTLTAPSGVEAQPVDLPVS----- 980
Qy 464 LPSLFHVQOPSGNEKQVTTISHSSTLTIOKCGKTMPTVNTIIPSGPP 514
Db 981 -----ILASP-----TTEQPTATVTIADSGGDVPGVTYVCSNPP 1017

RESULT 6
AMYH_YEAST
ID AMYL_YEAST STANDARD: PRT: 1367 AA.
AC P08640: P08640:
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1993 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucoamylase SI/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
  glucosidase) (1,4-alpha-D-glucan glucohydrolase).
GN ST11 OR ST12 OR MAL5 OR YIR019C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
  Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RA Churchill B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
  Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
  Gaultier S., Hamlyn N., Horsnell T.S., Hunt S., Jagals K., Jones M.,
  Lous E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
  Rajadream M.A., Riles L., Rowley N., Skelton J., Smith V.,
  Walsh S.V., Whitehead S.;
  Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RL [2]
RN SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX MEDLINE=87194600; PubMed=3106330;
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
  STA1."
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STA2 and SGA genes
  from Saccharomyces cerevisiae."
RL FEBS Lett. 239:179-184(1988).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
  glucose residues successively from non-reducing ends of the chains
  with release of beta-D-glucose.
CC -1- SIMILARITY: TO S.POMBE SPC215.13.
CC -1- SIMILARITY: SOME, TO S.POMBE SPC285.13C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation at
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.isb-sdb.ch/announce/
  or send an email to license@sdb.ch).
CC -----
DR EMBL: Z38061; CA86176.1;
DR EMBL: M16164; AAA35014.1;
DR EMBL: M16165; AAA35015.1;
DR EMBL: X13857; CAA32069.1;
DR PIR: B26877; B26877.

```

```

DR PIR: A26877; A26877.
DR PIR: S48478; S48478.
DR SGI: S0001458; MUC1.
KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
  KM Signal; Multigene family.
FT SIGNAL 1 21
FT CHAIN 22 1367
FT DOMAIN 210 1367
FT CARBOHYD 817
FT CARBOHYD 874
SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2BDB61AAAD CRC64;

Query Match 8.0%; Score 220; DB 1; Length 1367;
Best Local Similarity 22.5%; Pred. No. 5.5e-05;
Matches 131; Conservative 77; Mismatches 240; Indels 134; Gaps 22;

Qy 2 TLVTKVAPVSAVPKVGSGRLPAP-----QIVAKAPPTTIOIPANQLPQGVLIK 54
Db 618 TTESSAPVTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTE 677
Qy 55 SNGPILVSPQOT---VTRAETTSNITSRAVAPDQVTKICTVPSNSQLIKVAV-- 109
Db 678 SSSAPVTSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPV 737
Qy 110 -TVYKLAIGITVTT-----VPRSSVQS-----VAVPTSVTVTPGKPLNT- 152
Db 738 PTPSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTP 797
Qy 153 -----VTLKPSLSGASTPSNEPNLAENSAVAQINLSPMLNENVKCKNPLA 201
Db 798 SSSTTESSAPVPTPSSSNITSAPSSITPSSSTTESSAPV---PTPSSSTTESS-- 851
Qy 202 MLIKLACSGSQS-----PEMGQNVKLIQDLDAKIEEPTRLVYELKSSPPHLY 254
Db 852 --APVSSSTTESSAPVPTPSSSNITSAPSSITPSSSTTESSAPV---PTPSSSTTESS-- 904
Qy 255 PFLAKSVVALROLPLPSQSITQOCVQOOTSMDYIACITTVTTSPP---VTTTVSSQSE 311
Db 905 -----YPSQQT--ETVSSTTETTPKRTTVSVTPPTTITVTCSTGTN 949
Qy 312 KSIIIVGATAPRTV--SVQTLNPLAGPVGAKAVTLHSGVPAAGTGTAGC--LLQT 367
Db 950 SAGETTSKSPKRVTTTPTTTTSTTSTTTTTCSTGNSAGETTSKSPKRTTT 1009
Qy 368 SKPLVTSVANT---VTVSLQPEKPVVSGTAVTLSPAVT--FGETSGAIC----- 414
Db 1010 TVPCSTPSETASESTTSPPTTVTVSVTVVTVETSKRGELITTVTKNIPITY 1069
Qy 415 -----LPSVKPVVSCWCHICKPVIGTPOIKLADPGVLSQPAIPTGSSSKOLFSLF 468
Db 1070 LTTIAPPTPVTVTVTF-----TPPTTITTT---VCS-----TGTNS----- 1101
Qy 469 HVVQOPSGNEKQVTTISHSSTLTIOKCGKTMPTVNTIIPSGPP 510
Db 1102 -AGETTSKSPKRVTTTVCSTGT-----GEYTTAETTLVTVTA 1138

RESULT 7
MUC2_HUMAN
ID MUC2_HUMAN STANDARD: PRT: 5179 AA.
AC 002817; Q14878;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mucin 2 precursor (Intestinal mucin 2).
GN MUC2 OR SMUC.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;

```

RX MEDLINE=94132002; PubMed=8300571.
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to Prepro-von Willebrand factor";
RL J. Biol. Chem. 269:2440-2446(1994).
RN [2]
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=93016075; PubMed=1400449;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
RA Kim Y.S.;
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT both upstream and downstream of its central repetitive region.";
RL J. Biol. Chem. 267:21375-21383(1992).
RN [3]
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX MEDLINE=91358717; PubMed=188563;
RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA Petersen G.M., Kim Y.S.;
RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT and polymorphism.";
RL J. Clin. Invest. 88:1005-1013(1991).
CC -1- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
CC OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A
CC PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS
CC AGENTS AT MUCOSAL SURFACES.
CC -1- SUBUNIT: MULTIMERIC.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
CC BRONCHUS, CERVIX AND GALL BLADDER.
CC -1- PFM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC -1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND
CC VARIES AMONG DIFFERENT ALLELES.
CC -1- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT
CC OF SLIKORM HEMOCYTIN.
CC -1- SIMILARITY: CONTAINS 2 WFEC DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTEINE KNOT-LIKE (CTCK) DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L21998; AAB95295.1; -;
DR EMBL: M74027; AAA59875.1; -;
DR EMBL: M94131; AAA59163.1; -;
DR EMBL: M94132; AAA59164.1; -;
DR Genew: HGNC:7512; MUC2.
DR MIM: 158370; -;
DR InterPro: IPR000359; Cys_knot.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002400; GE_cysknot.
DR InterPro: IPR002919; TIL_Cysrich.
DR InterPro: IPR001007; VWF_C.
DR InterPro: IPR001846; VWF_D.
DR Pfam: PR00007; Cys_knot; 1.
DR Pfam: PR00093; VWC; 1.
DR Pfam: PR00094; VWD; 4.
DR Pfam: PR01826; TIL; 1.
DR PRINTS: PR00438; GFCYSKNOT.
DR SMART: SM00214; VWC; 2.
DR SMART: SM00216; VWD; 4.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS01185; CTCK_1; 1.
DR PROSITE: PS01225; CTCK_2; 1.
DR PROSITE: PS01208; VWF_C; 2.
DR GLYCOPTOLIN: Repeat; Signal.
FT SIGNAL 1 20 POTENTIAL.

FT CHAIN 21 5179 MUCIN 2.
FT DOMAIN 1401 1747 APPROXIMATE REPEATS.
FT REPEAT 1401 1416 1.
FT REPEAT 1417 1432 2.
FT REPEAT 1433 1448 3.
FT REPEAT 1449 1464 4.
FT REPEAT 1465 1471 5.
FT REPEAT 1472 1478 6.
FT REPEAT 1479 1494 7A.
FT REPEAT 1495 1517 7B.
FT REPEAT 1518 1533 8A.
FT REPEAT 1534 1556 8B.
FT REPEAT 1557 1572 9A.
FT REPEAT 1573 1596 9B.
FT REPEAT 1597 1612 10A.
FT REPEAT 1613 1635 10B.
FT REPEAT 1636 1651 11A.
FT REPEAT 1652 1675 11B.
FT REPEAT 1676 1683 12.
FT REPEAT 1684 1699 13.
FT REPEAT 1700 1715 14.
FT REPEAT 1716 1731 15.
FT REPEAT 1732 1747 16.
FT DOMAIN 4815 4886 WFEC 1.
FT DOMAIN 4924 4991 WFEC 2.
FT CTCK.
FT DISULFID 5075 5122 BY SIMILARITY.
FT DISULFID 5089 5136 BY SIMILARITY.
FT DISULFID 5098 5152 BY SIMILARITY.
FT DISULFID 5102 5154 BY SIMILARITY.
FT ? 5159 BY SIMILARITY.
FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 770 770 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 894 894 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1139 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1154 1154 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1215 1215 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1230 1230 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1246 1246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1266 1266 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1787 1787 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1820 1820 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4339 4339 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4351 4351 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4362 4362 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4373 4373 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4422 4422 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4438 4438 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4502 4502 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4616 4616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4627 4627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4752 4752 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4787 4787 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4881 4881 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4888 4888 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4955 4955 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4970 4970 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5019 5019 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5038 5038 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5069 5069 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1351 1351 H -> L (IN REF. 3).
FT CONFLICT 1412 1412 T -> S (IN REF. 3).
FT CONFLICT 1449 1449 L -> P (IN REF. 3).
FT CONFLICT 1504 1504 M -> T (IN REF. 3).
FT CONFLICT 4192 4192 G -> S (IN REF. 2).
SQ SEQUENCE 5179 AA; 540295 MW; 85CD7571FB9A5663 CRC64;

Query Match 7.9%; Score 219; DB 1; Length 5179;
Best Local Similarity 21.9%; Pred. No. 0.00031;
Matches 136; Conservative 63; Mismatches 227; Indels 194; Gaps 29;
QY 2 TLVTKVAVSAP-----PKVSSGPRLLPAPQIVAVKADTTTI-----QPPAN 43


```

FT CARBOHYD 3065 3065 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3144 3144 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3172 3172 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3288 3288 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3292 3292 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3782 3782 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4005 4005 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4136 4136 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4243 4243 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4254 4254 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4335 4335 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4376 4376 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4586 4586 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 5136 5136 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 5252 5252 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 5376 AA; 579908 MW; 0E4ADB77DE2A2620 CRC64;

Query Match 7.2%; Score 197.5; DB 1; Length 5376;
Best Local Similarity 23.2%; Pred. No. 0.0052;
Matches 146; Conservative 79; Mismatches 234; Indels 171; Gaps 33;

QY 8 APVSAB--PKVSSGPRLPAPQIVAVKAPNTTITQFPA---NLQLPFGVLLIKNSGGLM 61
DB 551 SPLPPTGPESEVPLPMQPTSPKATVTITETPTTEATITPTTETTVPE---VI 607
QY 62 LVSPOQTVAETTSNITSRAVPANPQYKICTVPSNSSLIKKVAVPYKLAQIGTT 121
DB 608 MSPKETSIPPEVT-----IPTEVITV-----SPEELISPEVTPVDVTAAYV 652
QY 122 VVTVT--PKPSY--QSVAVPTSVYTPGKPLNTVTLKPSST--GASTPENENLKA-E 177
DB 653 EATNASPEETSVPEVTITLEVTVSPEE--TTPPEVPIVLIDAPFAPEGETTLTEVP 710
QY 178 SAAVOINLSPTMLENKKCNFLAMLIKLAGSSGSPENQNKILVEQLDLAK---IEA 234
DB 711 TVPTEVTGVTETVTV-----SPE---ETSVPTEETISREVTTVSP 748
QY 235 EETTRK-----LYVELKSSQPHL-----VPFLKKSVALROLPLNPSOSTIQOQVOOTS- 283
DB 749 EETTVPEVPIVLEATASPTGEITLYTEVPVPEVTVGVTETVTVNSP-----EETSV 802
QY 284 --SDMVIATCTTV-----TTSPTVTTVSSSGSEKSIIVSGATAPRYSVOTLPLAGPVG 338
DB 803 PTEETISTEVTYTSPEETITLPTVEPTVSTEVTVNS-----PEETSVPEETI----- 849
QY 339 AKAGVTLHSVGPAAATGGTACTAGLL-----QTSKPLVTSVANTVTVNS-----IQPEK 388
DB 850 ----LTTLVTEVPTPEVTVGVTETVTVNSPEETSVPEETISTEVTYTSPEETITLPTEV 905
QY 389 PVSAGNAVLSLPVAFEGETSGAICLPYKP-----VSFCMDHICKPVIGTPQIK 441
DB 906 PTVS--TEVTNVSPPEET-----SVPEETITLITETTVSPPEETVPPIGTTLV-- 949
QY 442 LAQPGVLSOPAGI-----PTGSSSKOLFSLFHVQOPSNGEKOVTTISHSST----- 490
DB 950 ----PTEVLVPIVITVFPPIGETT--VPTEVPYVSTENTGVTETVTVFPPEETISIPREVA 1004
QY 491 -----LTIQKCGOKTQMPVN--TITPTSQPPE---ASTLKQITLPGNKILS 530
DB 1005 VLPASIPPEETTPTEETITTPPETITIPAEVTVPPASIPPEETASILEVTTTTPPETTT 1064
QY 531 ----LQASPOKNIKIKENVT-----SCFRDE 552
DB 1065 PTEVTVVPEKTTITPTEVTVPPASIPPEE 1094

RESULT 10
DAN4_YEAST
ID DAN4_YEAST STANDARD: PRT; 1161 AA.
AC P47179;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)

```

```

DE Cell wall protein DAN4 precursor.
GN DAN4 OR YJR151C OR J2223.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_Taxid=4932;
RN (1)
RP SEQUENCE FROM N.A.
RA Scarce T.?
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN (2)
RP REGULATION.
RX MEDLINE=21113168; PubMed=11160904;
RA Cohen B.D., Serfil O., Abramova N.E., Davies K.J., Lowry C.V.;
RT "Induction and repression of DAN1 and the family of anaerobic
RT mannoprotein genes in Saccharomyces cerevisiae occurs through a
RT complex array of regulatory sites.";
RL Nucleic Acids Res. 29:799-808(2001).
CC -1- FUNCTION: COMPONENT OF THE CELL WALL (By similarity).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SRP1 / TIR1 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC
DR EMBL; Z49651; CA89684.1; -.
DR SGD; S0003912; DAN4.
DR InterPro; IPR000992; SRP1_TIR1.
DR Pfam; PF00660; SRP1_TIR1.1.
DR PROSITE; PS00724; SRP1_TIR1.1.
KW Cell wall; Glycoprotein; Membrane; GPI-anchor; Signal.
FT SIGNAL 1 24
FT CHAIN 25 1146 CELL WALL PROTEIN DAN4.
FT PROPEP 1147 1161 REMOVED IN MATURE FORM (POTENTIAL).
FT LIPID 1146 1146 GPI-ANCHOR (POTENTIAL).
SQ SEQUENCE 1161 AA; 118358 MW; 7954C15069F0CA58 CRC64;

Query Match 7.1%; Score 195; DB 1; Length 1161;
Best Local Similarity 21.1%; Pred. No. 0.0011;
Matches 149; Conservative 100; Mismatches 242; Indels 216; Gaps 33;

QY 4 VTKVAPVSAAPKSSGPRLPAPQIVAVKAPNTTITQFPAQLPPTVLIKNSGFL--- 60
DB 369 VTSSAEPTTVSEVTSVEPTRSSQVTSAPPTVSEPTSVSE--PTRSSQVTSAPPTVTS 427
QY 61 -----MLVSPQQTVAETTSNT-----TSRAVPANPQYKICT---VNSSS 101
DB 428 EFTSSVEPTRSSQVTSAPPTVSEPTSVSEPTRSSQVTSAPPTVSEPTSVSEPTTRSS 487
QY 102 QLIKRVAVTPVKKLAQIGTTVTVTPKPSVQSVAAVPTVTPG--KPLNT--VTTLP 158
DB 488 QVNSAEPPTVSEF-----TSSVEPI--RSSQVTSAPPTVSEPTSVSEPTRSSQVTTTTP 542
QY 159 --SSLGAS--STPSNRP-----NKAENSAVOINLS-----PYMLENVKKCNFLANL 203
DB 543 VSSFSGTFSSETTSAPLPSKATTSAGSISNQIITSSBLIYSSVITSSSETPSSIEVL 602
QY 204. IKLAGSGSGSPENQNKILVEQLDLAKIEAE-----FTFKKLVLELKSSQPHLYVP 256
DB 603 TSSGISSVEP-----TSLVGPSSDESISSTSLATSTFTSSAVVSSSKAA-----DF 650
QY 257 LKRSVALROLPLNPSQFIQOQVQOTSMDVAVTATCTTTVTTSFVTVTT----- 304
DB 651 FTRSTVSASAKSDVSGNS-----TQSTTFEATPSTPLAVSTVTVTSSDVSAPNIPPS 702
QY 305 -VSSSGSEKSIIVSGAT--APRTVSQ-----TLNPLAGPVGAKAGVTVTLHSGVP 351

```


RA Glockner G., Scherer S., Schattevoy R., Borlight A., Weber J.,
MEDLINE=5501010; PubMed=9/99/93;
RA

RA Tsui L.C., Rosenthal A.;
 RT "Large-scale sequencing of two regions in human chromosome 7q22:
 RT analysis of 650 kb of genomic sequence around the EPO and CMT1 loci
 RT reveals 17 genes.";
 RL Genome Res. 8:1060-1073(1998).
 RN [2]
 RP SEQUENCE OF 2338-2700 FROM N.A.
 RC TISSUE-Testis; PubMed-9126492;
 RA MEDLINE-97271566; Garbers D.L.;
 RT Gao Z., Harumi T.,
 RT "Chromosome localization of the mouse zonadhesin gene and the human
 RT zonadhesin gene (ZAN).";
 RL Genomics 41:119-122(1997).
 CC -1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
 CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
 CC SIGNALING.
 CC -1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
 CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.
 CC -1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
 CC ZONA PELLUCIDA.
 CC -1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
 CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
 CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVUDUCTAL ISTHMUS.
 CC -1- DOMAIN: THE WVD DOMAINS 2 AND 3 MAY MEDIATE COVALENT
 CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
 CC -1- SIMILARITY: CONTAINS 3.5 MAM DOMAINS.
 CC -1- SIMILARITY: CONTAINS 4.5 WVD DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF053356; AAC78790.1; -;
 DR EMBL: U83191; AAC51208.1; -;
 DR GeneW: HGNC:12857; ZAN.
 DR MIM: 602372; -;
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000998; MAM_domain.
 DR InterPro: IPR002919; TIL_Cysrich.
 DR InterPro: IPR003328; Tila_Cysrich.
 DR InterPro: IPR001846; WVF_D.
 DR Pfam: PF00094; WVD; 4.
 DR Pfam: PF00629; MAM; 4.
 DR Pfam: PF01826; TIL; 5.
 DR Pfam: PF02345; Tila; 4.
 DR SMART: SM00216; WVD; 1.
 DR PROSITE: PS01186; EGF_2; 3.
 DR PROSITE: PS00740; MAM_1; 1.
 DR PROSITE: PS50060; MAM_2; 4.
 KW Glycoprotein; Transmembrane; Cell adhesion; Repeat.
 FT NON_TER 1
 FT DOMAIN <1 109 MAM 1.
 FT DOMAIN 112 136 MAM 2 (PARTIAL).
 FT DOMAIN 161 326 MAM 3.
 FT DOMAIN 322 446 MAM 4.
 FT DOMAIN 483 951 66 X HEPTAPEPTIDE REPEATS (APPROXIMATE)
 FT DOMAIN (MUCIN-LIKE DOMAIN).
 FT DOMAIN 953 1065 WVD 1 (PARTIAL).
 FT DOMAIN 1066 1454 WVD 2.
 FT DOMAIN 1455 1861 WVD 3.
 FT DOMAIN 1862 2292 WVD 4.
 FT DOMAIN 2293 2684 WVD 5.
 FT DOMAIN EGF-LIKE.
 FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 403 403 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1023 1023 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1099 1099 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1618 1618 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1737 1737 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1832 1832 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1878 1878 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2136 2136 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2505 2505 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2374 2379 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 2700 2700 NNOKMA -> RAGPGP (IN REF. 1).
 SQ SEQUENCE 2700 AA; 293013 MW; 80E60CC0B12277B1 CRC64;
 Query Match 7.1%; Score 194.5; Db 1; Length 2700;
 Best local Similarity 22.7%; Pred. No. 0.0033;
 Matches 138; Conservative 84; Mismatches 235; Indels 151; Gaps 33;
 QY 1 GLTVTRV-APVAPP-----RVSSGPRLPAPQIV-AVRAPPTTII 38
 DB 372 GTMLELLGSPASPPIPLMKRVGSQRPYQNTSVTVPSGHQPMQLIFKIGSGNTAVS 431
 QY 39 QPAPNQLPPGIVLINSGLMLVSPQCTVTRAETTSNTSRPVPANPQVYKICTVFN 98
 DB 432 VANGFTLINGCPVRV-LPELPVSPVSTGSETT-GLTENPTISTKRPVS----- 483
 QY 99 SSSQLIKVAVTPVKLAQIGTVTVTPKSSVQSAVAVTSVTVYVPGPLNTVTTLMP 158
 DB 484 -----TEKSVTEK-----PTVRK-----EKPTIPEKPTISTEK- -TIPSEK 522
 QY 159 SSLGASTSPSENPENLKAENSAVOINLSPTMLENVKCKNFLMLIKLCSGSPSEMGQ 218
 DB 523 -----NMPSEKPTISE-----KPTILEKPTIS-----EKPTIPEKPTI-- 559
 QY 219 NVKKIVBQLDAIEAEFTFRKIYELKSSPOHLVPLFKSVVALROLPNQSTI--Q 276
 DB 560 STEKPTVPEEPTPEETT--TYME-----EP-VIPEKPSIPEKPSIPEKPTISME 611
 QY 277 QCVQOTSMDVATCTTTVTSVTVTSVSSQSEKSIISG-----ATAPRVS-----V 327
 DB 612 ETIISTEKPTICEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTI 671
 QY 328 QTINPLAGP--VGAKAGVTVLHSVGPATGTTAGTGLQTSKPLVTSVANTVT--VS 383
 DB 672 PTEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTI-----ISTEPTPEETIISTEKS 726
 QY 384 LQPEKPVSTANTVLSLPAVTFBETSGAACLVSVPVVSFCMDHICKVIGI-----P 437
 DB 727 IPEKPTLPTEETTSVEETI--STEKLTIPMEKPTIS-----TEKPTIPEKPTIS 778
 QY 438 VQI-----KLAQPG-----PV-----LSQPAIGPTSSSKQLFSLFHVQOPSQGN 478
 DB 779 EKLTIPTEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIS-----TEKPTI 834
 QY 479 EK-----QVTTISHSSTLIQKCGQKTMVNTIIPTSQPPASILKQTLPGNKILSLQA 533
 DB 835 EKPTIPEETIIS-TEKLTIPTEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTI-----T 887
 QY 534 SPFOKNRI 541
 DB 888 IPEKPTI 895
 RESULT 13
 VGS0_HSVI1
 ID VGS0_HSVI1 STANDARD; PRT; 670 AA.
 AC Q00130;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Hypothetical gene 50 protein.
 GN 50.
 OS Ictalurid herpesvirus 1 (Chanel catfish virus) (CCV).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Ictalurid Herpes-like viruses.
 NC NCBL_TaxID=10401;
 RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN-Auburn 1;
RX MEDLINE=92087490; PubMed=1727613;
RA Davison A.J.;
RT "Channel catfish virus: a new type of herpesvirus.";
RL Virology 186:9-14(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M75136; AAA88153.1; -
DR PIR; F36791; F36791.
KW Hypothetical protein; Repeat.
FT REPEAT 143 158
FT REPEAT 171 186
FT REPEAT 200 214
FT REPEAT 215 233
FT REPEAT 234 252
FT REPEAT 253 268
FT REPEAT 279 293
FT REPEAT 294 309
FT REPEAT 320 334
FT REPEAT 335 349
FT REPEAT 362 376
FT REPEAT 377 391
FT REPEAT 392 406
FT REPEAT 407 421
FT REPEAT 422 436
FT REPEAT 437 452
FT REPEAT 464 477
FT REPEAT 478 493
FT REPEAT 504 517
FT REPEAT 518 531
FT REPEAT 532 545
FT REPEAT 546 559
FT REPEAT 560 573
FT REPEAT 574 587
FT REPEAT 588 601
FT REPEAT 602 615
FT REPEAT 616 629
SQ SEQUENCE 670 AA; 64174 MW; 2B64A781C519EBB4 CRC64;
Query Match 7.0%; Score 193.5; DB 1; Length 670;
Best Local Similarity 22.1%; Pred. No. 0.00069;
Matches 136; Conservative 60; Mismatches 276; Indels 143; Gaps 27;
QY 1 GTLVTKAAPSAPPK-----VSSGRLPAPQI--VAVKAPNT----- 35
DB 62 GDAVTGQPVREPSLSPTTPSLAHITISKMISLGR-PPPTTPGCVTTTPNDAPVP 120
QY 36 -----TTTQFPANLQLPPTGLIKNSGSLMLVSP-----OOTVTRA 72
DB 121 GSVHTTARVVDTTKOTPTTPATPAGA---NDTANITATPAGANDTANITATPAGA 176
QY 73 ETTSNITSRAVPANPQ--TYKICTVNSSQLKKVAVTVVKLAQIGTVVTVTP- KP 129
DB 177 NDTANITATPAGANDTANITATPAGANDTANITATPAGANDTANITATPAGANDTANITATPAMP 236
QY 130 SSVOSVAVPTSVTVTPGKPLN-----TVTLKPSISGASSTPENEPLKKAENSAVAQIN 184
DB 237 AGANDTANGTA-VTTTPAMPAGANDTANITATPT--GANDTANVTMAGATDIV---VT 290
QY 185 LSPMLLENVKKCKNPLAMLILACSGSSSPMGONVKKLVQQLDAKIEAEFTRLIVE 244
DB 291 TTPAMPAGANDTAN-ITATPAGANDTANVTMPAGATDVTTTPAMPAGANDTANVTKP 349
QY 245 LKS-----SPQHLVPLFKSVVALRQLLPRNSQSFIOCVQOTSSDVIATCTTVTSP 299

```

```

DB 350 AGSTDTVTTTPAMPAGATDVTTTPAMPATGAT---DTVVTTTPAMPAGATDVTTTP 406
QY 300 -----VYTTTVSSQSEKSIIVSGATAPRTVSQTLNPLAGVAKAGVYTLHSVG 350
DB 407 AKPAGANDTAVTTTPAMPAGANDVTVTTATPATPAGANDTAN-VKPPGATDVTTTATVK 465
QY 351 PPAATGGTTAGTG-----LLQTSKPLVTSVANTVTVSSLP-----EKPV-V 391
DB 466 PTGATGVTTTTPAKPTGANDTANVTKP--TGATGVTTTTPAKPTGANGTVVATAKPTGA 523
QY 392 SGTAVTSLPVAVTGEGTGA--AICLPSVKPVSEFCMDHICKPVIQIPVQIKIAQPGVYL 449
DB 524 TGVYTT-----TTAKPTGANGTVTTTAKPT-----GATGVTV 556
QY 450 SOPAGIPTGSSSKQLEFHVQOPSGENKQVTTISH--SSTLTIQKCGQKTMVPNTII 507
DB 557 TTTAK-PTGANG-----YTTTAKPAGANGTVTTTAKPAGANGTVTTTAKPAGANGTV 611
QY 508 PTSQPPASILKQIT 522
DB 612 TTTAKPAGANGTVT 626
RESULT 14
VGLX_HSVB STANDARD; PRT; 797 AA.
ID VGLX_HSVB
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Glycoprotein X precursor.
GN 71.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=31520;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9229556; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1.";
RL Virology 189:304-316(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M86664; AB02506.1; -
DR PIR; H36802; VGBEX1.
KW Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 22
FT CHAIN 23 797
FT DOMAIN 23 465
FT TRANSMEM 766 790
FT CARBOHYD 590 590
SQ SEQUENCE 797 AA; 80342 MW; 50C9ED9211F5E5B2 CRC64;
Query Match 6.9%; Score 189.5; DB 1; Length 797;
Best Local Similarity 18.9%; Pred. No. 0.0014;
Matches 118; Conservative 65; Mismatches 269; Indels 173; Gaps 15;
QY 1 GTLVTKAAPSAPPKVSSGRLPAPQIVAAKAPNTTITQFPANLQLPPTGLIKNSGSL 60
DB 45 GTTSSSPPTSPPTSSP-----PTSTHTSSPSTSTQ----- 79
QY 61 MLVSPQGVTPAETTSNITSRAVPANPQVYKICTVNSSQLKKVAVTVVKLAQIGT 120
DB 80 ---SSSTANISSAPSTASSSTISIPSTSTETTTTPASTTTPPTTAAFTAAAT---T 133

```

```

QY 121 TVYTVPKPSSVQSVAVPTSVTVTPGKPLNTVYTLKPSLSGASSTPSNEPNKAENSAA 180
DB 134 TAVTAASTSAETTTATATATSTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 193
QY 181 VOINLSPTMLENVKCKNFAMLIKACSGSGSPENQNVKKLVEOLDAKIAEEETPK 240
DB 194 TTTAATTTAATTTA-----ATTAAATTAA 218
QY 241 LVEELKSPQPHLVPLPKSVVALROLLENPSQSFIOCCVOOISSDMVIATCTTTVTSPV 300
DB 219 TTTAATTTA-----SATTAAATTAAATTAAATTAAATTAAATTAA 256
QY 301 VTTTSSSGSEKSIIVSGAATPRTVSQVLNPLAGPVGAKGVTLHVSQPTAATGTTA 360
DB 257 ATTGSPRTSGSISTTGAASSTPSASTATATATPTSTSAATTTST-----PTPSAATSA 311
QY 361 -GTGLQTSKPLV-TSVANVTTVTSLOPEKPVSGTAVTLSPAVT-----FGET 408
DB 312 ESTTEAPTSTPTDPTTPTSPATTTATSTPESTTVSASTSATTTAFTTESHTSPDSSTG 371
QY 409 SGA-----AICLPYKPVV-----SFCMDHICKPVIGT----- 436
DB 372 STAEPSSTPTLPSTPTSTPTDPTGSSASTESDSTDSSTVPTTGSTESSTESSSTTEASTN 431
QY 437 ----PVQIKLAQPGVLSQPAIGPTGSSSKOLFSLFHVQOPSGNEKQVTTTSHSTLT 492
DB 432 LGSSTESTETALEFTPDGNTTSGNTTSPSPRTSPSPADTQOTPDNGVSTQTTTINDHTAN 491
QY 493 IOK-----CGOKTMP-----VNTIITPSQFPASILKQI 521
DB 492 AQKHAGHHGRAGRGRRGSPGSGSHTRPHDRLPSPDDTYDDDTNHPNGANNIEIYVPL 551
QY 522 TLPGNKILSLQASPTOKNRKENV 546
DB 552 P-PDRPIELGVATLKRNMFEASCT 575

RESULT 15
XJH8_YEAST
ID XJH8_YEAST STANDARD; PRT; 881 AA.
AC P47033;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 89.2 kDa protein in SCPI60-SMC3 intergenic region.
GN YJ078C OR J1027.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C;
RX MEDLINE=96093911; PubMed=7483841;
RA Mioda T., Schaeff-Gerstenschlaeger I., Chaltatzis N., Baur A.,
RA Boles E., Fournier C., Schmitt S., Veltien C., Wilhelm N.,
RA Zimmermann F.K.;
RT "Sequence analysis of a 33.1 kb fragment from the left arm of
RT Saccharomyces cerevisiae chromosome X, including putative proteins
RT with leucine zippers, a fungal Zn(II)-Cys6 binuclear cluster domain
RT and a putative alpha 2-CB-alpha 2 binding site.";
RL Yeast 11:681-689(1995).
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
CC INSCTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: Z49353; CAA89370.1; -
DR EMBL: X83502; CAA58492.1; -
DR EMBL: X88851; CAA61314.1; -
DR HSSP: P04284; 1CFE.
DR SGD: S0003614; YJ078C.
DR InterPro: IPR001283; Allrgn_V5/7px1.
DR Pfam: PF00188; SCP_1.
DR PRINTS: PR00837; VSTPLXKE.
DR PRODOM: PD000542; Allrgn_V5/7px1; 1.
DR SMART: SM00198; SCP_1.
DR PROSITE: PS01009; SCP_AGS_PRL_SC7_1; 1.
DR PROSITE: PS01010; SCP_AGS_PRL_SC7_2; 1.
KW Hypothetical protein.
FT DOMAIN 166 379 ALA/SER/THR-RICH.
SQ SEQUENCE 881 AA; 89152 MW; 37DBAC660CA9D12A CRC64;

Query Match 6.7%; Score 184.5; DB 1; Length 881;
Best local similarity 21.3%; Pred. No. 0.0031;
Matches 130; Conservative 75; Mismatches 229; Indels 177; Gaps 24;

QY 47 PGCTVL-----IKNSGPLMLVSPQQTVAETTSNITS--RPAPANPQTV 91
DB 146 PPNYLGEFAEEVEPLISTVSSSS-----SSSTSTSDVSTISSIMPAV-AQGYTT 199
QY 92 KICTVNRSSQLKKKAVTPVK--KLAQICTVYTVTPK-----SSVQSAVPTSVT 143
DB 200 TVSSAASSSS--LKSTINPAKTATLTASSSTVYITSTESVGSSTVSSASSSVTSYAT 257
QY 144 ---VTPGKPLNTVYTLKPSLSGASSTPSNEPNLK-----AENSAVQILSPTME 191
DB 258 SSSTVVSQDASTTTTSSVATSSSTSSDPSTSTAASSSDPAASSMAAS--SSASTE 314
QY 192 NVKCKNFTA-----MLIKLACSGSGSPENQNVK-----KLVBQ 226
DB 315 NAASSSSAISTSSSNVSAPLSLTSTASTASRSVTSNVSVKFAVTVFSAQTTSSVA 374
QY 227 LLDKTEAEFEETRKLYVELKSSPPH--LYPFLKKSVALROLLRSSQ----FIQCQV 280
DB 375 SLSSVAADDIQGSTKEATSVSEHTSIYTSATNAQVATRLGSSRSSGAVSSAAS 434
QY 281 QTSDDNVIA-----ICTVTVTTSPTV--TVSSSGSEKSIIVSGA 319
DB 435 QSLNVSVIANVTQSVTSVSTAHTRKDATITVTSASESTSTSTNAASSSTEKNINSAA 494
QY 320 TAPRTVSVQTLNPLAGPVGAKAGVTLHVSQPTAAGTGTAGTLQTSKP----- 370
DB 495 T-----SSSIYSNSASVSGHGVTAAEYAIITSEOSALATSVPATNCSSIVK 541
QY 371 -----LYTSVANVTTVTSLOPEKPVSGTAVTLSPAV----- 403
DB 542 TTTLENSSTTTTATKSTTTTLATTTANNSTRAATVAVTID-PTLDPDONSASPTDNKHTS 600
QY 404 TFGETSGAALCLPSVPRVVSFCMDHICKPVIGTPVQIKLAQPGVL----- 449
DB 601 TVSSSTGA-SLDSLTTTTSISVSTTQLVSTCTSESDSPSAISTATTTESNLT 659
QY 450 -----SOPAGIPTGSSSKOLFSLFHVQOPSGNEKQVTTTSHSTLTQCGKTMP 502
DB 660 NTITACSDTSDSNPPTSAASSTDETATRTISTS-----CSDLNGASTGT-SLTLTSPMK 712
QY 503 VNTIITPSQFP 513
DB 713 TNTVPASSFP 723

```

Search completed: February 16, 2003, 21:56:01
 Job time : 25.6381 secs


```

Db 108 TTVOPTTLQSRPGVOPOLVLGSSAQPASLGTAFVOTGHPORVPGASTTSTATE--- 164
Qy 186 SPMLENVKKCKNFMLTKLACSSGSPGMCNVKLYEQLLDKIEAEFTRLKLYEL 245
Db 165 ---TMBENVKKCKNFSTLTKLSSGKOSTETANVNDLVONMLDKIEAEFTSRLYREL 221
Qy 246 KSSPOHVLPLFKSVVALROLLPNSOSFQOCVOQSSDMVATCTTWTTSPTVTTTV 305
Db 222 NSSPOYLPFLFKSLPALROLTPDSNAFTQSSQOQPP---ASQATLALIVLVSSSV 277
Qy 306 SSSSEKSIIVSGATAPRTVQTLNPLAGPVAKAGVVTLSVGPAAVGTAGTGLL 365
Db 278 ORTAGTAASVTSALQPVYISL-----TQPVGVGKQ 310
Qy 366 QTSKPLVTSVANTVTVTSLOPEKPVYSGTAVTSLPAVTFEGTSGAICLPSVKPVVSVFC 425
Db 311 APPETLVIO-----QPEKP-----GALIRPPOVTLTQT-----PMVALR 344
Qy 426 MDHICPVIGTPOVQIKLAPGVPVLSQAPAGIPTSSSKOLFSLFHVVOQPSGNEKQVTTI 485
Db 345 QPH-NRIMLTTPQOIQ----- 360
Qy 486 SHSSTLLTKCGCKTMTVNTIIPTSQPPASILKQITLPGNKIL---SLQASPTOKNRK 542
Db 361 -----NOLQPVYVVKPTVLTGTALSTVSAQAALAAQKNK 396
Qy 543 ENVTSCFRDE 552
Db 397 EPGGGSFRD 406

RESULT 2
ID 08T9E0 PRELIMINARY; PRT; 851 AA.
AC 08T9E0;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE SD047335P.
GN TAF110.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Broksstein P., Hong L., Agbayani A., Carlson J.,
RA Chang M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuncio J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rudin G.M., Cealiker S.;
RA Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY069807; AAL3952.1;
SQ SEQUENCE 851 AA; 92093 MW; 665B28B958C984C CRC64;

Query Match 8.5%; Score 234.5; DB 5; Length 851;
Best Local Similarity 24.1%; Pred. No. 2,2e-06;
Matches 140; Conservative 80; Mismatches 176; Indels 185; Gaps 29;

```

```

Db 227 CKRFNLNLEL---STREPKVEKNVRLTLOELVNAVNEPEEFCDRLERLINASPOCLIG 284
Qy 256 FLKRSVVALROL-----LPNSQSFIO-----OCVOQ 281
Db 285 FLKSLPLNLQALYTELVEIGIKPPQVHVLGLAGLSQQLPKIQADLRIGSQOTTIGQ 344
Qy 282 TSSDMVATCTTWTTSPTVTTTVSSSQSEKSIIVSGATAPRTVS---VOTLNLPLAGPVG 338
Db 345 TOYRMI---TPNALGTPTPTTGHITTSKOPN---IRLPTAPRLVNVGIGTQIP-SLOVP 398
Qy 339 AKAGVVTLSVGPAAVGTAGTGLQTSKPLVTSVANTVTVSLOPEKPVVSGTAVTL 398
Db 399 GQANVIOIR---GPOHLOLQTSVQIRATRP-----PNSVPTAN----- 436
Qy 399 SLPAVTFEGTSGAICLPSVKPVVSECMHICKPVIGTPOVQIKLAPGVPVLSQAPAGIPG 458
Db 437 KLTAHVVGOTQIKAI-TPSLHP-----PSLAISGGRPP 469
Qy 459 SSSKOLFSLFHVVOQPSGNEKQVTTISHSSTLLTKCGCKTMTVNTIIPTSQPPASTI 517
Db 470 TPTLSVLS-----TLNSAST-----TTLPIPS-LPTVHLPEALR 503
Qy 518 ---LKOITLPGN-----KILSLQA---SPTOKNRKIKENT 546
Db 504 ARPOMONSLNHSNHPDAKLVEIKAPSLAPPHMERINASTL 544

RESULT 3
ID 029071 PRELIMINARY; PRT; 528 AA.
AC 029071;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Gastric mucin (Fragment).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE-GASTRIC MUCOSA;
RX MEDLINE=95275264; PubMed=7755593;
RA Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
RA Lamont J.T.;
RT "Isolation and characterization of cDNA clones encoding pig gastric
RT mucin.";
RL Biochem. J. 308:89-96(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN-YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE-GASTRIC MUCOSA;
RX MEDLINE=94102478; PubMed=7506218;
RA Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
RA Lamont J.T.;
RT "Pig gastric mucin: isolation and characterization of a cDNA clone
RT with a novel tandem repeat.";
RL Gastroenterology 106:200-200(1994).
DR EMBL; U10281; AAC48526.1;
FT NON_TER 1 528
FT SEQUENCE 528 AA; 49907 MW; 0BF0F66792032EA CRC64;

Query Match 8.5%; Score 233.5; DB 6; Length 528;
Best Local Similarity 22.3%; Pred. No. 1,4e-06;
Matches 126; Conservative 93; Mismatches 196; Indels 149; Gaps 26;

```


FT NON_TER 1 1
 SQ SEQUENCE 1349 AA; 135600 MW; 4DC3C1544F1E5EBA CRC64;

Query Match 8.1%; Score 224.5; DB 4; Length 1349;
 Best Local Similarity 22.2%; Pred. No. 1.6e-05;
 Matches 143; Conservative 80; Mismatches 279; Indels 141; Gaps 26;

QY 2 TLVTKVAP---VSAPPKVSQP---RLPAPQIVAAKAPNTTTIOPE-ANLQLPPTGLI 53
 DB 725 TSTTAPASTASTASTASTASTASTASTASTASTASTASTASTASTASTASTAST 784
 QY 54 KSSSGPLMLVSP-QQVTRAEFTSNITSRAPVAPNPQYKICVNPSSSLIKKAVATPV 112
 DB 785 ---SSPTTSTTPPQSTTSSPTTSTTAP-----TSTTAPPTTST-----TSTPQ 828
 QY 113 KKLAIQITVTVTPPKPSV-----QSVAVPTSVTVTPGKPLVNTTLKPS 159
 DB 829 TSTSSAPSTSTTAPASTASTASTASTASTASTASTASTASTASTASTASTAST 888
 QY 160 SLGASSTPSNEPNL-----KAENSAVAOINLSP----- 187
 DB 889 S-GSGTTPSPVPPTTSTASVSKTSTSHVSKTTHSQPVTRHCPHCTWTKWFDVDEPSPG 947
 QY 188 -----TMEENKCKCNFLAMLKILACSGSQSPE-----MGQNVKVLVEQLDAKI 232
 DB 948 PHGDKETVNNIIRSEKICRPEETRLQCRASHPEVSIHLLGOVVOCSREGVLVCRN 1007
 QY 233 EAEETPKRL---VYELKSPQPHLVPLKKSVALRQLLPNSQSTLOOCVOQTSDDMTI 288
 DB 1008 QDQGGFKKCLANEVAVLCEPCKGCPVSTPYTA-----PSTPSGRANSPTOSTSMOK 1062
 QY 289 AICTTAVTTSPVTTVSSQSEKSIIVGATAPRTVSQVTLNPLAGVCAKAGVATLHS 348
 DB 1063 SRTTTLVTTSTSTPQSTTSTASTSTIPAST--PSTTSAPTSTSTASTSTASTPTHT 1121
 QY 349 V-GPT-----AATGTTAGTGLQTSKPLVTY-ANTVTTVSLOPKKPVSGAVPLSLP 401
 DB 1122 TSGPTTSTLAPTTSTTASTASTASTASTASTASTASTASTASTASTASTAST 1181
 QY 402 -----AVTGETSGAICLPYKPVVSCMDHICKPVIGTPVQIKLAOPGVLSPAGI 455
 DB 1182 QTSKTSAASTSTSGSG--TTPSPVPTTS--TTSASTSTSTASTSTSTSGG--TTPSPV 1236
 QY 456 P-----TSSSKQLFSLFHVQOPSGNEKOVTTISHSSTLTIOKCG 498
 DB 1237 PSTSITSAAFTSTTAPPTTSTTASTPTSSM--TSGPTTSPVPTTSTTASTSTTSGP 1293
 QY 499 KTMPVNTIIPTSOPPASILKQITLPGN-----KILSLQASPT 536
 DB 1294 GTTP--SPVPTTSTTASTPTTSTSGPSTPSVPPTTSTTASTPT 1334

RESULT 6

Q9N4S7

ID Q9N4S7 PRELIMINARY; PRT: 1079 AA.

AC Q9N4S7: 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
 DE Y51B11A.1 protein.
 GN Y51B11A.1
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=990613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium.";

RL Science 282:2012-2018(1998).

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Cotton M.;
 RT "The sequence of C. elegans cosmid Y51B11A."
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL:AC006797; ANF60743.1;
 DR Interpro:IPR002965; P_rich_extensn.
 DR PRINTS:PRO1217; PRICHEXTENS.
 SQ SEQUENCE 1079 AA; 110532 MW; 8BDDE3824CF80CA1 CRC64;

Query Match 8.1%; Score 222.5; DB 5; Length 1079;
 Best Local Similarity 23.0%; Pred. No. 1.6e-05;
 Matches 130; Conservative 66; Mismatches 225; Indels 145; Gaps 24;

QY 5 TKVAPVSA---PKVSSGRLPAPQIVAAKAPNTTTIOPEANLQLPPTGLV-----IKSN 56
 DB 399 TTTAPETSTEPSSTTP-----VQTTTAPETSTEPSSTTPVQTTTAPETST 454
 QY 57 SGPMLVSPQOVTTRA--ETTSTSRPAPVAPNPQYKICVNPSSSLIKKAVATPVK 114
 DB 455 EPPSSSTTPVQTTTAPETTS--TEPPSSSTTPVQTTTAPETSTEPSSTTPVQ- 511
 QY 115 LAQIGTTVTVTPKPSVQSVANVPTSVYTPCKPLNTVTLKPS-----LGAS 165
 DB 512 -----TTTAPETSTEPSSTTPVQTTTAPETSTEPSSTTPVQTTTAPET 565
 QY 166 TPSENPILKAENSAVAOINLSPMLENVKCKNFLAMLKILACSGSQSPMGQNVKIVE 225
 DB 566 TSTEP---PSSSTTPVQTTTAPETSTEPSSTTPVQTTTAPET----- 611
 QY 226 QLLDAKIEAEFTKRLVELKSSPOHLVPLKKSVALRQLLPNS-----OSFIQCVQ 280
 DB 612 -----TSTEP-----PSSSTTPVQTTTAPET 634
 QY 281 QTSDDVATCT--TTVTSPTVTTVSSQSEKSIIVGATAPRTVSQVTLNPLAGPV 337
 DB 635 TSTEPSSTTPVQTTTAPETSTEPSSTTPVQTTTAPETSTEP-----PS 688
 QY 338 GAKAGVTLHSVGP--TAATGTTAGTGLQTSKPLVTSVANTVTSLOPEKPVSGTAV 396
 DB 689 SSNTPVQTTTAPETSTEPSSTTPVQTT--TTTAPETST-----EPPSSSTTPV 740
 QY 397 ---TSLPVTGE-----TSGAICLPYKPVVSCMDHICKPVIGTPVQ-IR 441
 DB 741 QTTTITAPETSTEPSSTTPVQTTTITAPETSTEPSST-----SNTPVQTTT 790
 QY 442 LAQPGVLSOPAGIPGSSSKQLFSLFHV-----VOQPSGNEK-OVTTISHSSTLTIOK 495
 DB 791 TTPPETSTEP---PSSSTTPVQTTTITAPETSTEPSSTTPVQTTTITAPETSTEP 847
 QY 496 CGQKMPVNTI---PTSOFPAS 516
 DB 848 PSSSTSPVQTTTITAPETSTEPSST 873

RESULT 7

Q9OY35

ID Q9OY35 PRELIMINARY; PRT: 1979 AA.

AC Q9OY35: 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
 DE Host cell factor CI (Fragment).
 GN ABCD1 OR HCFC1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Platzer M., Brenner V., Reichwald K., Wiene T., Oksche A.,
RT Rosenthal A.;
*Comparative sequence analysis of the mouse L1cam locus and the
RT corresponding region of human Xq28.*;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF133093; AAF22156.1;
DR MGD: MGI:105942; Hcfc1.
DR MGD: MGI:1349215; Abcd1.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF01344; Kelch; 4.
DR SMART: SM00060; FN3; 1.
FT NON_TER 1
SQ SEQUENCE 1979 AA; 203436 MW; ABE98C5DDE688A9 CRC64;

Query Match 8.0%; Score 222; DB 11; Length 1979;
Best Local Similarity 23.2%; Pred. No. 3.5e-05;
Matches 140; Conservative 71; Mismatches 232; Indels 160; Gaps 29;

QY 4 VTKAPVSAAPKVSSEPRP---PAPQIVAVKAPNTTIOFPANL-----QLPGETVL-- 52
DB 417 VPAVLKVTGPQATTTGTPLVTRMPASQ--AGKAPYVTV-SLPASVAMVVPVTSAGCTG 473
QY 53 -----IKNSGRLMLVSPQOTVTRAEFTSNITSAPVAPANQYKI 93
DB 474 NPQSGMAALAAATAATOKIPSSAPVLSVPAGT---IVKTVAVTGGTTTLEPATKV 529
QY 94 CTVP-----NSSOLIKKVA--TPVKLAIQIGTVTVTPKPSVSQAVAPTSTVTVTP 146
DB 530 ASSVWVSNPATRLKLTAAQVGTSSAANTSRPIITVAKSGTV-IVAQAQVTVTVV 588
QY 147 GKPLNTVTLK-PSGL-GASSTPSNEPNL-----KAENSAVOINLSPTMLENKKCN 198
DB 589 GGVTTKTTLVKSPIVSPPGSAIISNLGKMSVQTKPVQTSAVTQASGTGPTQIIQTKG 648
QY 199 FL--AMLIKLAGSGSSQSP-----EMGQNVKKLVEQLDLAKIEAEFTKKLYELKSS 248
DB 649 PLRPGTILKLTVSADGKPTTITTTQASGAGTKPTILGI-----SS 689
QY 249 PQPLVLPFLKSSVALROLPLNSQSFIOQVOQTSMDVIACTTCTTVTT-----S 298
DB 690 VSPSTT---KPGTTTIKTIPMSAIITQAGATVTSPPGKSPITITTKWTSGTGAPA 746
QY 299 PVTV-----TVSSSQSKSIIVSGA-----TAPRT-----VSQTLNPLAGPV 337
DB 747 KIITAVPKRIAGHQOQGVTVLKGAPGQPGTILRTVPMGCVRLVPTVASVAVKPAVTVTL 806
QY 338 GAK--AGVTLHSVGPPTAATGTTAGTGLQTSKPLVTSVA--NTVTVSLQPEKPVVSG 393
DB 807 VKRTTGTGTTIGTGYTST--SLAGAGHSTSLATPITTLGTLATLSSQ---VNP 860
QY 394 TAVTLSLPATVFGTSGAICLPKVPVSPCMHDICKPVGITPVQIKL-AQPGVLSQP 452
DB 861 TAITVSAAGTTLTAAGGLTPTITMPV-----SQPLQVULITAPSSVEAQP 907
QY 453 A-GIPGSSSSKQLSLFHVVOQPSGNGEKQVTTISHSTLTIOKCGQKTMVNTIIPPSQ 511
DB 908 VHDLPVS-----ILASP-----TTIDPATVITADSGQDVQPGFTVLVCS 948
QY 512 FPP 514
DB 949 NPP 951

DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE EMSY protein.
GN CILORF30.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Hughes-Davies L.;
RT "EMSY is amplified in breast cancer and displays a BRCA2 dependent DNA
RT damage response.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ430203; CAD22881.1;
SQ SEQUENCE 1322 AA; 141467 MW; 7F8C95E8BA0FC9F0 CRC64;

Query Match 8.0%; Score 219.5; DB 4; Length 1322;
Best Local Similarity 22.7%; Pred. No. 3.1e-05;
Matches 144; Conservative 88; Mismatches 230; Indels 171; Gaps 30;

QY 5 TKVAPVSAAPKVSSEPRPAPQIVAVKAPNTTIOFPAN--LQPLPG-TVLIKNSGRLM 61
DB 149 TTSPTSTP-----VPSGIAVTKSPRPAS--PASNMYVLPSSGVYKVSQSCDE 197
QY 62 LVSPQOTVTRAEFTSNITSRAPVAPNPQTV---KICIVPNSS-----SOLIKKVA----- 108
DB 198 DEKPKK---KRTTNSSSSPVYKPKAVPVSTTTVPVSGSKMSNMOSINSLPP 254
QY 109 -VTPVK-----KLAQIGT---VTTVPKSSQSV-----AVPTSVTV 144
DB 255 HNSPVKIFKTPSTQTTNTTQKVIITVTSPTSTVPPILSKSHYAAVTKVPTSVYSLA 314
QY 145 TPGLKNTVTLKPSL-----GASSTPSNEPNL-----LKAENSA 180
DB 315 TTQKP-PVYITASQSLVSNSSGSSSTSPSPITPTVAVTAVVSTPSVMSVTAQGVST 373
QY 181 VOINSPMLENVKCKNFMFLIKLAGSGSSQSPMGQNVKKLVEQLDLAKIEA-EETFR 239
DB 374 SAIKMASTRLSPKSLVAPQIIL-----AQFQHQQSRK--QQLYVOQOQTQOQVAAQ 425
QY 240 KLYELKSSPQPLVLPFLKSSVALROLPLNSQSFIOQVOQTSMDVIACTT---TV 295
DB 426 PSPVSHQOQPOQSPLPKIPKPTIQKQ-----ESGVKIITQOVQSKILPKVTVTLPTS 480
QY 296 TTSPTVTTVSSSQSKSIIVSGAIPRTVSVQTLNPLAGPVGAAGVTV----- 345
DB 481 SNSPLMVVSSMGALMTTKLVTTPTGTQATYTRPVPSPISGRMAATPGAAIYVKTSGSII 540
QY 346 -----LHSVG-----PRAATGTTAGTGLQTSKPLVTSVANTV---TVSLOPEKPV 391
DB 541 TVPKSLATLGGKTISSNIVSGTTIKITIPRTSPKSNVIVQKTKGKGTIQLGPKANV 600
QY 392 S-----GTAVTLSPATVFGTSGAICLPKVPVSPCMHDICKPV-IGTPVQIK--- 441
DB 601 TTLNAGGEKTIQVTP-----TGAKPAIILATRPITKMI---VTPQGISGTVQPAKI 651
QY 442 -----LAQGPVLSQPAIGIPGSSSKQLSLFHVVOQS----- 475
DB 652 IPTKIVYQOQKTOYLKPKPVTFOATVV--SEQTRQVTV--ETLQASRVAEAGNSSIQ 707
QY 476 -GNGEKQVTTISHSTLTIOKCGQKTMVNTII 507
DB 708 EGKEPQNTDSSSSSTESSQSDSQSFVAVHI 740

RESULT 9
ID 061191
AC 061191;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

RESULT 8
ID 08TE50
AC 08TE50;
PRELIMINARY;
PRT: 1322 AA.

01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Transcription factor C1 (HCF).
GN HCF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Kristie T.M., Dasher R.;
RT "cdnas encoding the mouse homolog of the human transcription factor C1
(HCF).";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U03925; AAB0163.1; -.
DR MGD; MGI:105942; Hcf1.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR001798; Kelch.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF01344; Kelch; 5.
DR SMART; SM00060; FN3; 1.
SQ SEQUENCE 2045 AA; 210537 MW; B89CC2FDA35B969F CRC64;

Query Match 7.9%; Score 217; DB 11; Length 2045;
Best Local Similarity 23.5%; Pred. No. 7.4e-05;
Matches 141; Conservative 76; Mismatches 224; Indels 158; Gaps 31;
QY 7 VAPVSAPPKVSQPRRL--PAQIVAVKAPNTTIOIPANLQIPGTVLIRKNSGPMLV 63
DB 486 VLKVTGQATGTPPLVIMRPAQ--AGKAPVTV-SLPASVVR--VVPQSAQGTIVGS 539
QY 64 SPQQT-----VTRAETTSNT--TSRP--AVPAN-----PQIVKICTVP 97
DB 540 NPOMSGMAALAAATAATQKIPSSAPTMMSVPAGTTIVKTVAVTPGTTLPATVAVKVASP 599
QY 98 -----NSSOLIKKAV--TPVKKLAQIGTVTVTPKPSVOSVAVPTSVYTPPKPL 150
DB 600 VMSNPATRMKLTAAAGVGSVSSAANTSTRPIITVAKSGTV--TVAQAQVYTTVVGCVT 658
QY 151 NTVTTLK-PSL-GASSTPSNEPNL-----KAENSAVOINLSPTMLENKKCKNFL-- 200
DB 659 KITLVKSPISVPGSALINLCKVMSVQTKPVQTSVAVGQASTGTVQIIRKGPLA 718
QY 201 AMLIKLACSGSQSP-----EMQNVKKLVLEQLDAKIEAEETRKLVLEKSSQPH 252
DB 719 GTILKLVTSADGKPTTIITTOASGAGTKPTIIGI-----SSVSPS 759
QY 253 LVPLKKSVALRQLPNSQSFIOQCVQTSMDVATCTTYYT-----SPVVT 302
DB 760 TT---KPGTTTIKTPMSALITQAGATGVTSPPGKISPTIITTKVMTSGTGAPAKIIT 816
QY 303 -----TVSSSQSEKSIIVSGA--TAPRT-----VSQVTLNPLAGPVGAK- 340
DB 817 AVPKIATGHQGOQGVTVLKGAPGPGTILRTVPMGCVRLVTPVYSAVAPATVTLVVGK 876
QY 341 -AGVTVLHSGPTPAAGTGTAGTGLQTSKPLVTSVA--NTVTVVSLQPKPVVSGTAVT 397
DB 877 TTVGTVLTGTVTGVST--SLAGAGHSTASLAPITTLGTIATLSQ---VINPFAIT 930
QY 398 LSLPVTFGTSGAALCLPVSFSCMDHICKPVIGTVOIKL-AQPGVLSQPA-GI 455
DB 931 VSAAGTTLTAAGGLTPTTITMOPV-----SPTQVYLLTAAPSGVEAQAQVNDL 977
QY 456 PTGSSSKQLSLFHVVOQPSGNEKQVTTISHSSTLTIOCKGOKTNPVNTIIPTSQPP 514
DB 978 PVS-----ILASP-----TTQGPATVYTIADSGGDVQPGVTVLVCNMP 1017

RESULT 10
Q9QWH2 PRELIMINARY; PRT; 2045 AA.
AC Q9QWH2
DT 01-MAY-2000 (Tremblrel. 13, Created)

01-MAY-2000 (Tremblrel. 13, last sequence update)
DE Transcription factor C1 (HCF).
GN HCF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Kristie T.M., Dasher R.;
RT "cdnas encoding the mouse homolog of the human transcription factor C1
(HCF).";
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U08082; AAD09225.1; -.
DR MGD; MGI:105942; Hcf1.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR001798; Kelch.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF01344; Kelch; 5.
DR SMART; SM00060; FN3; 1.
SQ SEQUENCE 2045 AA; 210546 MW; 7AD3BPCD78EABA9C CRC64;

Query Match 7.8%; Score 215; DB 11; Length 2045;
Best Local Similarity 23.5%; Pred. No. 9.8e-05;
Matches 141; Conservative 75; Mismatches 225; Indels 158; Gaps 31;
QY 7 VAPVSAPPKVSQPRRL--PAQIVAVKAPNTTIOIPANLQIPGTVLIRKNSGPMLV 63
DB 486 VLKVTGQATGTPPLVIMRPAQ--AGKAPVTV-SLPASVVR--VVPQSAQGTIVGS 539
QY 64 SPQQT-----VTRAETTSNT--TSRP--AVPAN-----PQIVKICTVP 97
DB 540 NPOMSGMAALAAATAATQKIPSSAPTMMSVPAGTTIVKTVAVTPGTTLPATVAVKVASP 599
QY 98 -----NSSOLIKKAV--TPVKKLAQIGTVTVTPKPSVOSVAVPTSVYTPPKPL 150
DB 600 VMSNPATRMKLTAAAGVGSVSSAANTSTRPIITVAKSGTV--TVAQAQVYTTVVGCVT 658
QY 151 NTVTTLK-PSL-GASSTPSNEPNL-----KAENSAVOINLSPTMLENKKCKNFL-- 200
DB 659 KITLVKSPISVPGSALINLCKVMSVQTKPVQTSVAVGQASTGTVQIIRKGPLA 718
QY 201 AMLIKLACSGSQSP-----EMQNVKKLVLEQLDAKIEAEETRKLVLEKSSQPH 252
DB 719 GTILKLVTSADGKPTTIITTOASGAGTKPTIIGI-----SSVSPS 759
QY 253 LVPLKKSVALRQLPNSQSFIOQCVQTSMDVATCTTYYT-----SPVVT 302
DB 760 TT---KPGTTTIKTPMSALITQAGATGVTSPPGKISPTIITTKVMTSGTGAPAKIIT 816
QY 303 -----TVSSSQSEKSIIVSGA--TAPRT-----VSQVTLNPLAGPVGAK- 340
DB 817 AVPKIATGHQGOQGVTVLKGAPGPGTILRTVPMGCVRLVTPVYSAVAPATVTLVVGK 876
QY 341 -AGVTVLHSGPTPAAGTGTAGTGLQTSKPLVTSVA--NTVTVVSLQPKPVVSGTAVT 397
DB 877 TTVGTVLTGTVTGVST--SLAGAGHSTASLAPITTLGTIATLSQ---VINPFAIT 930
QY 398 LSLPVTFGTSGAALCLPVSFSCMDHICKPVIGTVOIKL-AQPGVLSQPA-GI 455
DB 931 VSAAGTTLTAAGGLTPTTITMOPV-----SPTQVYLLTAAPSGVEAQAQVNDL 977
QY 456 PTGSSSKQLSLFHVVOQPSGNEKQVTTISHSSTLTIOCKGOKTNPVNTIIPTSQPP 514
DB 978 PVS-----ILASP-----TTQGPATVYTIADSGGDVQPGVTVLVCNMP 1017

RESULT 11
Q76894 PRELIMINARY; PRT; 1795 AA.
AC Q76894
DT 01-NOV-1998 (Tremblrel. 08, Created)

```

; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aljun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922,217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-217-1068

Query Match 7.9%; Score 219; DB 10; Length 5179;
Best Local Similarity 21.9%; Pred. No. 5.8e-06;
Matches 136; Conservative 63; Mismatches 227; Indels 194; Gaps 29;

OY 2 TLVTAVAPASAP-----PKVSSGRLPAPQIVAKAPMTT-----QFPA 43
DB 3765 TTTTIVTPPTPTGQTPTTPTTITTTTTPPTPTPTPTPTPTPTPTPTPTG 3824
OY 44 LQLPPTVLINSNSGPLMLVSPQOTVRAETTSNTS-----RPAVPANQTVKICT 95
DB 3825 TQTPPTPTTITTTT---VTPPTPTGQPTPTPTTITTTTTPPTPTPTPTPTPT 3881
OY 96 ---VPNSSQILKAVATPVKKLAQIGTIVTVTPKSSVQS -AAVP-TSVVTVTP--- 146
DB 3882 TTTTTPPTPTPTGQTPTTPTTPTT---TTTTPPTPTPTGQTPTTPTPTPTPTPT 3936
OY 147 -----GKPLNTVTTTKPS-SLAGSTSPNEPMLKENSAAVOINSPMLLENVKCK 197
DB 3937 PTGQTPTTPTTPTTITTTTTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 3983
OY 198 NFLAMLILACSGSSPPEMGONVKRLVEQLDAKIEAEFTKRLVLELAKSSFOPLHVPPL 257
DB 3984 -----TGTQTP-----TTPITTTTTPPTP----- 4004
OY 258 KKSVALROLPLNSQSFIQCVQOTSDMVATCT-----TNTSPVVTTVS 306
DB 4005 -----TPYG-----TQTPPTPTTITTTTTPPTPTPTGQTPTTPTTITTTTV 4046
OY 307 SSQSEKSIIVSGATAPRTVSVQTLNPLAGPVGAK---AGVTLHSVPTATGCT--- 358
DB 4047 TPTPTPTGQPTPTPTTPTTPTT-TVPPTPTGQTPPTTPTTITTTTTPPTPTGQPT 4105
OY 359 ---TAGIGLQTSPL-----VTSVANTVTVLSQP-----EKPVSGTAVTISL 400
DB 4106 TPTTPTTTPPTPTPTPTGQTPTTPTTPTTITTTTTPPTPTPTPTPTPTPTTPTT 4164
OY 401 PAVTGETSGAICLPSVKPVVSCFMDHICKPVICTPQIKLACGPVLSPAGIPTGSS 460
DB 4165 PPTPTGT-----PTTTPPT-----TTTTPPT-----PTPTGTQ-TGPTPTHS 4202
OY 461 SKQFLSLHVVQOPSGENKQVTTTSHSSTLTQKCGKTMPEVNTIIP---TSQFPAS 516
DB 4203 TAPIREL--TTSNPPESTPQTSRSTSSPLT-----ESTLSTLPALEMTKLAPST 4255
OY 517 ILKQITLPGNKLTLSQASPT 536
DB 4256 PTAPTTSGHILSPPTST 4275

RESULT 5
US-09-833-263-1068
; Sequence 1068, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Aljun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Stolk, John A.

```


QY 9 PVSAPPKSSGPRLPAPQIVAVKAPNTTTIQPANIQLPBGVILKNSGPMILVS--PQ 66
1841 PVTTP-----SDSIPRIPPLPSVTA-----AKLSPVASGCIHQSPPTVTETMR 1887
QY 67 QVTRAEETSNITSRAVAPNPOTVIGIVPNSSSOLIKVANTP--YKLAIGTIVT 124
1888 QEEPRQOSTPS-----PALPBDTKASDDVSSSTL--RKILMDPKYVSATYSTSVTT 1939
QY 125 TVPKPSSVSAVNPVSVTVTPCKPLNTVYTLKPSISGASSTPSNPNLKAENSAVQIN 184
1940 AIAEPYAPAPCLHEAPPPVDSKKPLEKTA--PVTNNSEIQAESVLAADKEKAPV- 1996
QY 185 LSPFMLENKCKCNFLAMLIKACSGSPBQGNKYLVEQLDKIAEEFTRLIYE 244
1997 IAPKITSVISR-----MVSIDLENSOKITLAKPAPQTLGLVSA----- 2036
QY 245 LKSSPPQPHVFLKSVLRQLRPPNSQSFIOQCVOGTSMDVIACTTIVTSP--VYT 302
2037 LVLGVNVSILP-----VNLKGPVKGSVTLKSLVSTPAPVNLKGPVNLVLPVNLVLT 2091
QY 303 TVVSSQSEKSIIVSGATAPRVSVOTLNPPLAGPVGAKAGVYTLHSGPTAAGTTAGT 362
2092 TPVNAI-----VGTVNAAPGVY-----NAAASAVNATASAVTVTAAGVTAAGTAT 2140
QY 363 GLLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAVTSLPAVTFGETSGAICLPYKPVY 422
2141 GTV-----TMAGAVIAPSTRK--- 2155
QY 423 SFCMDHICKPVIGTPVQIKLAOPG--PVL--SQPAGIPTGSSSKQLSLFHVYQDPGNGNE 479
2156 -----CQORASANENSRF--HPSMPVQIDRDADAGSGGLR-----VITSEB--- 2196
QY 480 KOVTTISHSSTLILKCGOKTM--PVNTIIPTSQFPASILIKQITLPGNKILSLQASPTOK 538
2197 --VLLISYS-----GQKTEGPQRIKISQIPPASAMD-----IEHQGS--VSK 2236
QY 539 NRK-ENWTS 547
2237 SQVKPDSVTA 2246
Db

RESULT 9
US-09-862-027-40
; Sequence 40, Application US/09862027
; Patent No. US20020142428A1
; GENERAL INFORMATION:
; APPLICANT: Hodge, Martin R.
; TITLE OF INVENTION: No. US20020142428A1el kinases and uses thereof
; FILE REFERENCE: 35800/234862
; CURRENT APPLICATION NUMBER: US/09/862, 027
; CURRENT FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 09/345, 473
; PRIOR FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 40
; LENGTH: 1601
; TYPE: PRT
; ORGANISM: C. elegans
US-09-862-027-40

Query Match 6.4%; Score 176.5; DB 10; Length 1601;
Best Local Similarity 20.6%; Pred. No. 0.0082;
Matches 123; Conservative 93; Mismatches 244; Indels 137; Gaps 28;

QY 5 TKVAPVAPPKVSSGPRLPAPQIVAVKAPNTTTIQPANIQLPBGVILKNSGPMILVS 64
514 TPIPPAPAPPH--SSAQOQIIPPLSTQ-----TSAEIQSAAQO-----SVPTMTIA 559
QY 65 --PQGVTAETSNITSRAVAPNPOTVIGIVPNSSSOLIKVANTVYKKAIGTIV 122
560 NIPAMPTSAQOPVLIS--PTSAAVFPVPTTMIHVPKPSSEIPVONVAT-----AAVPAANN 613
Db

QY 123 VITVPKSSVQSAVAPV-----SVTVTPCKPLNTVYTL-----KBSLIGASS---- 165
614 VPPSPAPFKETDIQTPTLQNTVPRITISFDASGLVINTPASIASPSAPATDVASTTAP 673
QY 166 --TPSNENPKAENSAVQIULSPTMLENVKCKCNFLAMLIKACSSQ-----SPBQGN 219
674 VITPAPPTTTTDCGAAASTTTEENKERRKSNRKVMMEL--LGDDESINFALVSCRIDPS 732
QY 220 VKLVEQL-----LDKIAEEFTRLIYELKSSPPQPHVAPVY-----KKS-V-VALR 265
733 HKSVTQFAPGTDPCTINAKLAEOLKLVHVIYEAQGEVYQILNSGKKGVTGKLA 792
QY 266 QLI--PNSQS--FIQOCVQOQTSMDVIACTTIVTSPVYTT--TVSSQSEKSIIVSGAT 320
793 TVLDPNSTEPTTITVAWPKDSS--AATASNTKPKIEIKTEPTPTBASQDPNNVQV---T 846
QY 321 APTVSVQTLNPLAGPVGAKAGVYTLHSGPTAAGTTAGTGTG-----LQTSKPL 371
847 NVKRVQESNAESVQSTIPRPGIIVM--SPVQTSAPPPTGAALKPSHFQYTKSADP 903
QY 372 VTSVANTVTVTSLOPEKPVVSGTAVTSLPAVTFGETSGAICLPYKPVVSCMDHICK 431
904 ATRISSISATV--IPVIAATPTNI-----TSEPVYIQPTAQYIT----- 943
QY 432 PVIGTPVQIKLAOPGVY--LSQPAGIPTGSSSKQLSLFHVYQDPGNGNEQVTTISHS 488
944 -----HLATPPSPVSHLSNSNS--PSATTHSNMSI-----QSTTSVPG 980
Db

QY 489 STILIKCGCKTMVNTIIPTSQFPASILIKQITL-----GNKILSLQASPTOK 537
981 RRTVQPVQSAESGSISSISTPHPEPTAITSCPVPVSVVSGTINLEVAPO 1037
Db

RESULT 10
US-09-801-368-110
; Sequence 110, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Milne, Todd
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801, 368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487, 558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160, 587
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 110
; LENGTH: 1075
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-110

Query Match 6.3%; Score 174; DB 10; Length 1075;
Best Local Similarity 21.8%; Pred. No. 0.00072;
Matches 139; Conservative 80; Mismatches 244; Indels 174; Gaps 29;

QY 2 TLVTKVAPVAPPKVSSGPRLPAPQIVAVKAPN-----TTTIOFPANLQIPGTVILKSN 56
Db

DF 01-NOV-1999 (TREMblrel. 12, last sequence update).
 DT 01-JUN-2002 (TREMblrel. 21, last annotation update).
 DE EG:5667.1 protein.
 GN EG:5667.1 OR CG14796.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 NC NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananidis P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Adell J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA DePalos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,
 RA Palazolo R., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
 RA Styraskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-T., Wasserman D.A., Weinstock G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "the genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Cadieu E., Dreaano S., Lelaure V., Motlier S., Galibert F.,
 RT "Sequencing the distal x chromosome of Drosophila melanogaster";
 RL submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Benos P.;
 RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AEO03421; AAF45644.1; -;
 DR EMBL; AL01028; CAI19845.2; -;
 DR FLYBase; FBgn0025390; EG:5667.1.
 DR InterPro; IPR002557; Chitin_bind_PerA.
 DR InterPro; IPR002965; P_rich_extensn.
 DR Pfam; PF01607; CBM_14_2.
 DR PRINTS; PR01217; PRICHEXTENS.
 DR SMART; SM00494; ChEBD2; 2.
 SO SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;
 Query Match 7.8%; Score 214.5; DB 5; Length 1795;
 Best Local Similarity 20.2%; Pred. No. 9e-05; Mismatches 130; Conservative 86; Mismatches 254; Indels 173; Gaps 22;

QY 9 PVSAPPKVSGPRLPAPQIVAVKAPNTTIOFPANLQI-----PGTVLIKNSNG 58
 DB PTSSGKPTTPK-PSRTPTTKTKVTQTQITTTPLRSSSTETSTQPPPTTTPQPTT 758
 QY 59 PLMLVSPQOYTRATTTNINISRAVPANPQTVKICVTPNSSOLIKVAATVPKKLQI 118
 DB TLLVTPTKSTTTTTEKPTTSSPK-PTTQOKTTSTAPNT----KVAITQKETPT 812
 QY 119 GTT-----VTTVPKPS----VQSAVPTSVTVTPGKPLNT 152
 DB QSTSTTITRTKTTNNPEPTSEKPTSTTKRPSSTTKTSVASTKTISSKPTTE 872
 QY 153 VTTKPSLSGASSPSPSPNKAENSAAVQINLSPMLNKKCNFLAMLIKLAGSSQ 212
 DB KSTENPTNSVKTSLTSSQRA-----TSTSEPTKQNTITTPPKPTTKTS 922
 QY 213 SPEMQNKVLVEQLDLAK--TEAEFTKRLVELKSSPQHLVFLKSVVALQOLPN 270
 DB TQEAFTSTOKSTVTITTKKATESSPLTTLSTEENTPPKP-----LRTTPT 970
 QY 271 SOSFIOCCVOOTSDMTATCTTVTSPVVTYTTSSOSEKSIIVGA--TAPR--TVS 326
 DB 971 TTS-----VTAITRITTTTISESTETSTQKRSKSTPTTTRTPVTVI 1017
 QY 327 VQTLNPLAGPVGAKAGVYTLHSV-----PTATGGTA-GTGLQTSKPL 371
 DB 1018 VSTQNPPT--TTSKSTVTITTPNPSPTQRTTTRQPTSTASTSIGTRITPTNP 1075
 QY 372 VTSVANTVTVYSLOPEK-----VSGTAVTSLPVTGEGSGAICLPSPVVSF 424
 DB 1076 QNSTSTLTITVTRPCPDPTSDKNTACTOELQOVNLE-----LOSPOKQDF 1128
 QY 425 CWDHICKPVIGT-----PYQIKIAPGPVLSQAPGIFGSSSKQLSF 468
 DB 1129 THTRHTALTSGRNLTGQEPVDYMDAPSSAENSGCATTKAKPTMCTILAAHLQKLF 1188
 QY 469 HNV-----QOPSGNEKOYTT-----ISHSTLTITQRCG 497
 DB 1189 HIIPTPSREHAPQRPSSQSSQSRGVTIAQMARHNLATSKPFIASHRLSIQOLA 1248
 QY 498 ---OKTMEVNTI-----PTSGPPASIIKQITLGNKTL 529
 DB 1249 STQKRSIPKTLVHTNTKEPDESYDSEYDSEYDDEBDNEVL 1291
 RESULT 12
 Q8TDH7 PRELIMINARY; PRT; 1029 AA.
 AC Q8TDH7;
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
 DE Intestinal membrane mucin MUC17 (Fragment).
 GN MUC17.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NC NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21845432; PubMed=11855812;
 RA Gum J.R., Jr., Crawley S.C., Hicks J.W., Szymkowski D.E., Kim Y.S.,
 RT "MUC17, a novel membrane-anchored mucin";
 RL Biochem. Biophys. Res. Commun. 291:466-475(2002).
 DR EMBL; AF430017; AAL89737.1; -;
 FT NON_TER 1
 SO SEQUENCE 1029 AA; 108320 MW; 5E1AD19E4BF948E3 CRC64;
 Query Match 7.8%; Score 214; DB 4; Length 1029;
 Best Local Similarity 21.9%; Pred. No. 4.9e-05; Mismatches 135; Conservative 76; Mismatches 259; Indels 146; Gaps 24;

```

OY 4 YPKAVASAPKYSGRPLPAQIVAVKANNTTIOFPALQLPQGVILKNSGMLMV 63
DB 137 VDSSTVSTSTQSNSTP--TPPEVITLPMSTPEVSTPLTI-MVSTSTVITSEAGTAST 193
OY 64 SPOQVTRAFETTSNITSRPAVPANPQVKICTVPSNSSOLIKKVAATPVKLAIGTTVV 123
DB 194 LPVDISTPVITISQVSSSPVTPRG-TTMTPLWT-PSGEGTTLTMTPVSTTAVTSSEGSTLS 251
OY 124 TTVPKRSSVQSVAVPTSVVTPTPGKPLNTVY-----TLKPSLGASSTPSNEPLKAENS 178
DB 252 T-----PSVATSTPVTSTETATLSSATLSDTMTSMVSMPEISTLGTITLAVSTPTVTRPES 307
OY 179 AAVQINLSPMLLENVKKCKFLAMLKLACSGSQSPMGQNVKLVQLDAATEAEFT 228
DB 308 SPSST--PSV-----YTSMSMTTASGSSSP-----TLEGT 337
OY 239 RKLVELKSSPOPLVPELFKKSVALROLLPNSQSFQOCVQOTSSDMVATGTTVYTS 298
DB 338 TTPMSTTSEBSLTLTVL---ISPISVMSPEASSTLSTPGDTSTPLSTAGSFSIP 394
OY 299 PVVTTVSSSQSEKSIIVSGATAPRTVSQVQLNPLAGPVGAKAGVYTLMSVGPTAATGCT 358
DB 395 AEVTTIRISTERS-----TFLTLTVST-----TL---PTSPFGAS 429
OY 359 TACTGLQTSKPLVTSVANTVTVSLOPEKPVVSGTAVTSLPAVTPGETSGAICLPV 418
DB 430 IASTPLDSTSTFTPS---TDTASTPTLPV---ATTISVITIEGSGPTTIFIPST 480
OY 419 KPVVSEFCMD-----HICKPVI-----GTPVOIKLA--- 443
DB 481 -PVTSTADVFPAATGAVSPVITSTELNPNSTSSSTTTSFTKTEFTTPAMTAAPLT 539
OY 444 -----QPGVLSQAPGIPITGSSSKOLFLEHVQO-----PSGNEKQVTTISHST 490
DB 540 YVTMSTAPSTPRTISKGCSTASATLSTPHSTSTVTRPVYPSSESSSPSTTSHTP 599
OY 491 LTIQKCGKTMVNTIIPISQFPPA-SILKQITLPGKILSL-----QASPTQKNRIKEN 544
DB 600 PTPPASHSTPTTSASSTVNPENAVTMTTRKPKSTRTTSFTPTVTTAAVPT-NTTIKNS 658
OY 545 VNS-----CFRD 551
DB 659 PTSTPTVPRTTTCFGD 674

RESULT 13
P91365 PRELIMINARY; PRT; 2232 AA.
AC P91365;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE K06A9.1 protein.
GN K06A9.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-BRISTOL N2;
RC Giesel C., Gattung S.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS (A AND B) MAY BE PRODUCED BY
CC OF THE A FORM.
CC ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS THAT
DR EMBL; U80846; AAC70889.1; -
DR EMBL; U80846; AAC70890.1; -
KM Alternative splicing.
FT VARSPLIC 842 866
FT VARSPLIC 870 1051
VASSPAPSTSONPSTSSGSMI -> LATTSAPKPSVT
CLEMND (IN ISOFORM B).
PYPSOSTSPESSSTPSGSPGTLTSTSPSPSOSTTICST
CGSTSPGISTSEEMTSQSTQTPGSTGSTVTPQSTVSDST

```

```

FT SSGSTVTVGSTEBSSTPSTSONTPSTSSGSMSTQTPQ
FT SSGSTSPVESSTSGATSSSGSPGTLTSLSPSPSSSTIGS
FT SSGSTSPVESSTISQSGTE -> KEIDOTAINTKTYENAL
FT LVASKLNNEILTYIDNFGYSAGLNDHQYPPIDDYNGKIS
FT VPPIDGTDDDIDLDKVDYKSLATADMPVPADONCMFTI
FT SAAPDEYGGTTIKSTVTVPEYVYGVLVGAKSIRPLSDIK
FT NIVITNNTMDNRASAVASKLELPLFA (IN ISOFORM
FT B).
SQ SEQUENCE 2232 AA; 213840 MM; 08D69FA638E14CC8 CRC64;

Query Match 7.7%; Score 212.5; DB 5; Length 2232;
Best Local Similarity 21.9%; Pred. No. 0.00015;
Matches 138; Conservative 73; Mismatches 247; Indels 173; Gaps 24;

OY 26 QIVAVKANNTTIOFPALQLPQGVILKNSGMLVSPQVTRAFETTSNITSRPAVP 85
DB 801 QSVSTNPGSTVTR-----PSTVSGSTSSGTVTVGSTEASTSSSVASSPAPSTIS 852
OY 86 ANPQVTKICTVPSNSSOLIKKVAATPVKLAIGTTVTVTPRPSVQSVAVPTSVVTVT 145
DB 853 QNPV-----PSTSS-----GSSMITQSPYPS-QSTSPVESSSTPS 886
OY 146 PKPMLNTVTLKPS-----SIASSPNSM-----EPNL 173
DB 887 PGSPGTLTSTSPSPSGSTTIGTQGSTSPGISTSEEMTSQGSTQTPGSGTGTVPST 946
OY 174 KAENSAA--VOINLS-----PTMLENVKKCKNFLAMLIKLAGSGS---OSPENGON 219
DB 947 VSDSTSSGTVTVGSTEAGSSSPSTSONTPNS-----SSGSMSTQTPQSSQS 996
OY 220 VKKLVEQLLAKIEAEFTKRLVYELKSSPOPHL-----VPL 257
DB 997 TSP-VESSTSGAATSSSGSPGTLTSLSPSPSSSTIGSSGSTSPVYSTISOGSTETPGS 1055
OY 258 KKSVALRQL-----PN-SQFQOCVQOTSSDMVIA 289
DB 1056 TGSVTVTPSVVSGSASGSTATMGSTASSTSGSSSTSPSPSOSTBSSTGATSSPSSCG 1115
OY 290 TCTTTVTSPVTVTVVSSQSEKSIIVSGATAPRTVSQVQLNPLAGPVGAKAGVYTLHSV 349
DB 1116 TLTLSISPSQSSTIGSSGSTSPVYSTSGDWTSGSGTQIP-----GSTGTVTPQST 1170
OY 350 GPTAAT-----GCTTACTGLQTSKPLVTSVANTVTVSLOPEKPVVSGTAV---TL 388
DB 1171 GSGSTSTSGEITTSQGSTQTPRSSLSTSPALSTSQOVSSTNS--PGSTVTPQSTVTRGSTS 1228
OY 399 SLPAVTFGETSG-----AALCLPSYKPVVSEFCMD-HICKPVIATPVQIKLAQPGVLS 450
DB 1229 SGSTVTTTGSTEGSTSGSSSATSLSSSPVSTISQSPNPTSGSSTPTPNPSSGTSFVVS 1288
OY 451 QPAGIPTGSSSKOLFS-LFHVVOQP---SGGNEKQVTTISHSSTLTITQKCGORTMPVN-T 505
DB 1289 TTGEMTSHSGSTQTPRTIGSTVQTPSVSGSNSGSGTVTIGSBSATSGSFPSTSPSSIS 1348
OY 506 IIPTSOPPASILKQITLPGNKILSLQASPT 536
DB 1349 PVPTSPPTSTPTTASST-SGSTISDVSSVST 1378

RESULT 14
O9VGC9 PRELIMINARY; PRT; 842 AA.
AC O9VGC9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CPN protein.
GN CPN OR CG4795.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;

```


[1]
RN SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RA MEDLINE-20196006: PubMed-10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abdl J.F., Agbayan A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.T., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokya D., Butchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostel R.C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Slier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Sylkstra R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissensbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gdbbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003694; AAF54755.1; -
DR FLYbase; FBgn0010218; Cpn.
DR InterPro; IPR002965; P_flyc-extensn.
DR PRINTS; PRO1217; PRICHEXTENSN.
SQ SEQUENCE 842 AA; 82242 MM; D71E531327EF8501 CRC64;
Query Match 7.7%; Score 212; DB 5; Length 842;
Best Local Similarity 22.1%; Pred. No. 5.2e-05;
Matches 125; Conservative 81; Mismatches 236; Indels 124; Gaps 24;
QY 1 GTLVTKV-APVSAP-----PKVSSGRLPAP-----QIVAVKAPNTTIIOPANLQ 45
DB 4 GTTSPVSAIPAAVPTPSAIPAPOVVSIPAAPAAPIAVTPAPPTTASVOPAVVT 63
QY 46 LPPTGLIKNSGPMILVSPQOTVTRAEFTSNITSRPAVPANPQTVKICTVPNS--SSOLI 104
DB 64 IPAPAPIAASVAPASVAP--PVVAAPTPP-----AASPVSTPPVANAQIVANASAPVA 116
QY 105 KKVAVT-----PVKKLAOIGTT--VTTVPKSSVQSAVAPTSVTVTPGKPLNTVTTLK 157
DB 117 PPVATTPPVAPIPVAPVIAIPVPAASAPPAATPVVSP--VIATPPVVPANTTV--- 171
QY 158 PSSLGASTPSPNEPMLKAKNSAAVOINISPTMLE-----NKKCKKNFLAMLI 204
DB 172 PYAAIPVAAVPAVPAVPAVLAIPAAPVAAETPAPPAEIPVATIPPC--VAAPLI 228
QY 205 KLASGSSPPEMGQVKKLVEQLDAKTEAEFEFTKLYELKSSPQHLVFLKKSVAL 264
DB 229 -----PEVSVAATKPLAAAEPPVVAAPATEIPVVAAPAAASHVSAVPAVERAAYVA- 278
QY 265 ROLLPNOSFFIOQCVQOISSDMVIAITCTTIVTTSPPVTTTVSSSSEKSIIVSGATAPRT 324

DB 279 ----PVASAS-----TEPPVAAATLTTPAPT-PALAPVVAESQVAANTVATPTPTP-A 324
QY 325 VSVOTLNPAGVYKAGACAGVTLHSVGPAAATGCTAGTGLDTSPLVTSVNTVTVSL 384
DB 325 PEPETIAP-----PVAAETPEVASAVAAETTPPVPPVA-----AE 360
QY 385 QPEKPVSGTAVTLSPAVTGETSGAAICLPSPVPVVS-----FCMDHICK 431
DB 361 SIPAPVVAATTPPARUL-AVTDPDVASAV--PELPPVIAISPVSVAETPVDLAPVLP 417
QY 432 PYGTIPVOIKIAQPPVPVLSQPAIGTPTGSSSKOLEFLHVQOQSGCNEKQVTTISHSTL 491
DB 418 PVAAPPVVAVVAEETPEPPAPASAVTTIAALDIPVAPVIAAPSDAPAE-----APSAAA 472
QY 492 TIQKCGOKTMPNTTIIPTSQPPASI 517
DB 473 PT-----VSTPPTTASVETTPAPAAV 494
RESULT 15
ID 095045 PRELIMINARY; PRT; 864 AA.
AC 095045;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DR 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE GH08002P.
GN CPN OR CG4795.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyridae; Drosophilidae; Drosophila.
OX Nchi_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayan A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Flise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celiker S.,
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY058315; AALJ3544.1; -
DR FLYbase; FBgn0010218; Cpn.
SQ SEQUENCE 864 AA; 84565 MM; E2B96CFB0CB9802 CRC64;
Query Match 7.7%; Score 212; DB 5; Length 864;
Best Local Similarity 22.1%; Pred. No. 5.3e-05;
Matches 125; Conservative 81; Mismatches 236; Indels 124; Gaps 24;
QY 1 GTLVTKV-APVSAP-----PKVSSGRLPAP-----QIVAVKAPNTTIIOPANLQ 45
DB 4 GTTSPVSAIPAAVPTPSAIPAPOVVSIPAAPAAPIAVTPAPPTTASVOPAVVT 63
QY 46 LPPTGLIKNSGPMILVSPQOTVTRAEFTSNITSRPAVPANPQTVKICTVPNS--SSOLI 104
DB 64 IPAPAPIAASVAPASVAP--PVVAAPTPP-----AASPVSTPPVANAQIVANASAPVA 116
QY 105 KKVAVT-----PVKKLAOIGTT--VTTVPKSSVQSAVAPTSVTVTPGKPLNTVTTLK 157
DB 117 PPVATTPPVAPIPVAPVIAIPVPAASAPPAATPVVSP--VIATPPVVPANTTV--- 171
QY 158 PSSLGASTPSPNEPMLKAKNSAAVOINISPTMLE-----NKKCKKNFLAMLI 204
DB 172 PYAAIPVAAVPAVPAVPAVLAIPAAPVAAETPAPPAEIPVATIPPC--VAAPLI 228
QY 205 KLASGSSPPEMGQVKKLVEQLDAKTEAEFEFTKLYELKSSPQHLVFLKKSVAL 264
DB 229 -----PEVSVAATKPLAAAEPPVVAAPATEIPVVAAPAAASHVSAVPAVERAAYVA- 278
QY 265 ROLLPNOSFFIOQCVQOISSDMVIAITCTTIVTTSPPVTTTVSSSSEKSIIVSGATAPRT 324

```

Db 279 ---PVSAS-----TEPPVAATILTTAET--PALAPVAESQVAANTVVAATPTTP-A 324
QY 325 VSVQTLNPLAAPVAKKAGVUTLH5VGPPLAATGGTTAGTGLQTSKPLVTSVANTVTTVSL 384
Db 325 PEPETIA-----PVAAETPEVASVAEAETTPPVPEVA-----AE 360
QY 385 QPEKRPVSGTAVTL5LPAVTFEGTSGAICLP5VKPVVS-----FCMDHICK 431
Db 361 S1PAPVATTVPAFL-ATVPDPTASAV--PELPPIA5SPVPGAAETPPVDLAPVLP 417
QY 432 PVIGTPVOIKLAPQPVLSQPA5IPTGSSSKOLE5LFHVQOPS6GNEKQVTTISH5STL 491
Db 418 PVAAEPVAVVAAEETPETPAPASAPVITIALUDIPEDVAPVIAAD5DAPAE-----AP5AA 472
QY 492 T1QKCGKTMVNTIIP5QPPASI 517
Db 473 PI-----V5TPPTTASVPETIAPPA5V 494

```

Search completed: February 16, 2003, 21:58:53
Job time : 50.77 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:40:36 ; Search time 0.766595 Seconds
(without alignments)
1738.213 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_452

Perfect score: 53

Sequence: 1 AQPGLVLSQP 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: A_Geneseq_101002.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	53	100.0	801	19	AAW31494	Human hTAFII105 pr
2	53	100.0	801	22	ABG09468	Novel human diago
3	53	100.0	852	21	AA157279	Transcription fact
4	44	83.1	136	22	AAU45315	Protonibacterium
5	43	81.1	71	22	AAW06483	Human foetal prote
6	42	79.2	348	22	ABG14007	Novel human diago
7	41	77.4	353	21	AAW29893	Arabidopsis thalia
8	41	77.4	362	22	AAU00449	Maize disease resi
9	41	77.4	365	21	AAW29892	Arabidopsis thalia
10	41	77.4	397	21	AAW29891	Arabidopsis thalia

11	41	77.4	592	20	AAW97359	A 6-Transmembrane
12	40	75.5	63	23	ABP11105	Human ORFX protein
13	40	75.5	113	22	ABB03730	Human musculoskele
14	40	75.5	199	22	ABG20476	Novel human diago
15	39	73.6	24	22	AAE03611	Human hTAD protein
16	39	73.6	24	22	AAE82363	Human immune/hema
17	39	73.6	46	22	AAW83153	Human leucine-rich
18	39	73.6	794	22	AAE03600	Human 5' EST secre
19	38	71.7	794	22	AAE82352	Human polypeptide
20	38	71.7	55	20	AAV12001	Human breast tumou
21	38	71.7	136	22	AAO00050	Human protein sequ
22	38	71.7	290	20	AAV48475	Human protein sequ
23	38	71.7	397	22	AAW5506	Human protein sequ
24	38	71.7	554	22	AAW55064	Human protein sequ
25	38	71.7	570	22	AAW52921	Human protein sequ
26	38	71.7	608	21	AAV1537	Human secreted pro
27	38	71.7	694	20	AAV17829	Human secreted pro
28	38	71.7	694	21	AAW01330	Human PRO354 polyp
29	38	71.7	694	21	AAW01330	Human PRO354 polyp
30	38	71.7	694	21	AAW01330	Human PRO354 polyp
31	38	71.7	694	22	AAU29053	Human secreted pro
32	38	71.7	716	21	AAV91686	Human membrane or
33	38	71.7	875	22	AAW88346	Human growth-assoc
34	38	71.7	942	21	AAV32169	Human growth-assoc
35	38	71.7	942	21	AAW32169	Human growth-assoc
36	37	69.8	111	22	AAW59220	Human brain expres
37	37	69.8	111	22	AAW71755	Human bone marrow
38	37	69.8	111	22	AAW32047	Peptide #6084 enco
39	37	69.8	111	23	ABG41569	Human peptide enco
40	37	69.8	186	22	AAW65740	Fen-like protein.
41	37	69.8	194	20	AAV07465	Mouse T610q23.3 ge
42	37	69.8	220	22	AAW94488	Human protein sequ
43	37	69.8	349	19	AAW83021	Mouse cysteine-rich
44	37	69.8	349	19	AAW83021	A murine Dickkopf
45	37	69.8	349	21	AAV92069	Murine DKF-3. Mus

ALIGNMENTS

RESULT 1
AAW31494
ID AAW31494 standard; Protein; 801 AA.
AC AAW31494:
XX 28-APR-1998 (first entry)
DE Human hTAFII105 protein.
KW hTAFII105; transcription factor; TFIID; transcriptional activation;
KW antibodies; diagnosis; therapy; biopharmaceutical industry.
XX Homo sapiens.
OS US5710025-A.
PN 20-JAN-1998.
PD 02-OCT-1996; 96US-0725012.
PR 02-OCT-1996; 96US-0725012.
XX (REGC) UNIV CALIFORNIA.
XX Dikstein R, Tjian R;
XX WPI: 1998-109818/10.
XX N-PSDB: AAV02872.
XX DNA encoding human tata-binding protein associated factor - for
XX producing recombinant protein

XX Claim 1: Col 17-22; 12pp; English.
PS
XX This cDNA sequence represents a human tata-binding protein associated
CC factor, htraf1105, isolated from Daudi cell nuclear extracts. Tightly
CC associated subunits (TAF's) are components of the transcription factor
CC TFIID and are thought to mediate transcriptional activation. This encoded
CC protein may be produced recombinantly from transformed host cells or
CC purified from human cells. htraf1105 specific binding agents such as
CC specific antibodies could be used for diagnosis (e.g. genetic
CC hybridisation screens for htraf1105 transcripts), therapy (e.g. gene
CC therapy to modulate htraf1105 gene expression) and in the
CC biopharmaceutical industry (e.g. as immunogens, reagents for isolating B
CC cell specific activators or other transcriptional regulators).
XX
SQ Sequence 801 AA;
Query Match 100.0%; Score 53; DB 19; Length 801;
Best Local Similarity 100.0%; Pred. No. 0.94;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AOPGPVLSQP 10
Db 443 AOPGPVLSQP 452
|||||
RESULT 2
ABG09468
ID ABG09468 standard; Protein; 801 AA.
XX
AC ABG09468;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #9459.
XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
XX
DR N-PSDB: AAS73655.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20: SEQ ID No 39827; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 801 AA;
Query Match 100.0%; Score 53; DB 22; Length 801;
Best Local Similarity 100.0%; Pred. No. 0.94;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AOPGPVLSQP 10
Db 443 AOPGPVLSQP 452
|||||
RESULT 3
AAV57279
ID AAV57279 standard; Protein; 852 AA.
XX
AC AAV57279;
XX
DT 06-JUN-2000 (first entry)
XX
DE Transcription factor subunit TAFII105 polypeptide.
XX
XX TATA box-binding protein associated factor II 105; TAFII105; cancer;
KM transcription factor; apoptosis; cytostatic; immunosuppressive;
XX antinflammatory; virucide; antibacterial.
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FH Misc-difference 802
FT /label= Glx
FT /note="encoded by TGA"
FT Misc-difference 834
FT /label= Glx
FT /note="encoded by TAA"
XX
PN WO200012699-A1.
XX
XX
PD 09-MAR-2000.
XX
PF 25-AUG-1999; 99WO-IL00464.
XX
PR 27-AUG-1998; 98IL-0125971.
XX
PA (YEDA) YEDA RES & DEV CO LTD.
XX
PI Dikstein R, Yamit-hezi A;
XX
XX
DR WPI: 2000-256640/22.
XX
DR N-PSDB: AAZ90465.
XX
PT Polypeptide encoding TATA box binding protein associated factor II 105
PT useful for treating e.g. cancers and inducing apoptosis has a dominant
PT negative effect on the normal biological activity of the binding
PT protein -
XX
PS Claim 7: Fig 2; 48pp; English.
XX
XX This represents a polypeptide comprising a (modified) fragment (I) of
CC a TATA box-binding protein associated factor II 105 (TAFII105). A
CC pharmaceutical composition comprising (I) or the polynucleotide or an

CC inhibitor or antagonist of (I) is useful for treating cancers and
CC inducing apoptosis in pathological cells. The composition is also useful
CC for treating autoimmune diseases, inflammatory processes and viral or
CC bacterial infections.

XX
SQ Sequence 852 AA;

Query Match 100.0%; Score 53; DB 21; Length 852;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AOPGPVLSQP 10
|||
DB 443 AOPGPVLSQP 452

RESULT 4
AAU45315
ID AAU45315 standard; Protein; 136 AA.

XX
AC AAU45315;

XX
DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #6211.

XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.

XX
OS Propionibacterium acnes.

XX
PN WO200181581-A2.

XX
PD 01-NOV-2001.

XX
PF 20-APR-2001; 2001WO-US12865.

XX
PR 21-APR-2000; 2000US-199047P.

XX
PR 02-JUN-2000; 2000US-208841P.

XX
PR 07-JUL-2000; 2000US-216747P.

XX
PA (CORI-) CORIXA CORP.

XX
PI Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;

XX
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX
DR N-PSDB; AAS59525.

XX
DR WPI; 2001-616774/71.

XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for

XX
PT vaccinating against and diagnosing infections, especially useful for

XX
PT treating acne vulgaris -

XX
PS Example 1; SEQ ID NO 6510; 1069pp; English.

XX
PS Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins in the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by

CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 136 AA;

Query Match 83.0%; Score 44; DB 22; Length 136;
Best Local Similarity 80.0%; Pred. No. 5.5;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AOPGPVLSQP 10
|||
DB 61 AOPGPVLSQP 70

RESULT 5
AAM06483
ID AAM06483 standard; Protein; 71 AA.

XX
AC AAM06483;

XX
DT 05-OCT-2001 (first entry)

DE Human foetal protein, SEQ ID NO: 214.

XX
KW Human: foetal protein; cytosolic; immunosuppressive; immunostimulant;
KW neurotropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;
KW gene therapy; antitense therapy; cancer; immune disorder;
KW nervous system disorder; osteoporosis; thrombolytic disorder;
XX
OS Homo sapiens.

XX
PN WO200155339-A2.

XX
PD 02-AUG-2001.

XX
PF 25-JAN-2001; 2001WO-US02723.

XX
PR 25-JAN-2000; 2000US-0491404.

XX
PR 15-SEP-2000; 2000US-0663870.

XX
PR 06-NOV-2000; 2000US-0707351.

XX
PA (HYSE-) HYSEQ INC.

XX
PI Yeung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, Tang YT;

XX
PI Liu C, Asundi V, Zhou P, Werhman T;

XX
DR N-PSDB; AAH94158.

XX
DR WPI; 2001-46571/50.

XX
PT Novel fetal proteins useful for the treatment and diagnosis of diseases

XX
PT associated with dysfunction of the protein e.g. cancers, immune

XX
PT disorders and inflammation -

XX
PS Claim 10; Page 258; 715pp; English.

XX
PS The invention relates to novel foetal polypeptides encoded by
CC polynucleotides comprising one of 477 sequences fully defined in the
CC specification. The foetal polynucleotides and polypeptides are
CC useful in the treatment and diagnosis of diseases such as cancers,
CC immune disorders, growth disorders (e.g. osteoporosis), thrombolytic
CC disorders, nervous system disorders and inflammation. The present
CC sequence is a polypeptide encoded by a cDNA assembled using
CC an expressed sequence tag (EST) found to be expressed in human
CC foetal tissue cDNA libraries.

XX
SQ Sequence 71 AA;

Query Match 81.1%; Score 43; DB 22; Length 71;
Best Local Similarity 70.0%; Pred. No. 4.3;

Matches 7: Conservative 3: Mismatches 0: Indels 0: Gaps 0:

OY 1 AOPGPVLSOP 10
Db 40 SRGPVLSOP 49

RESULT 6

ABG14007
ID ABG14007 standard; Protein; 348 AA.

AC ABG14007;

DE 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #13998.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR N-PSDB; AAS78194.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

PS Claim 20; SEQ ID NO 44366; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 348 AA;

Query Match 79.2%; Score 42; DB 22; Length 348;
Best Local Similarity 80.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AOPGPVLSOP 10
Db 99 AOPGPVLSOP 108

RESULT 7

AC AG29893
ID AG29893 standard; Protein; 353 AA.

AC AG29893;

DE 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 35642.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 18-MAY-1999; 99US-0134370.

PR 19-MAY-1999; 99US-0134768.

PR 20-MAY-1999; 99US-0134941.

PR 21-MAY-1999; 99US-0135124.

PR 24-MAY-1999; 99US-0135353.

PR 25-MAY-1999; 99US-0135629.

PR 27-MAY-1999; 99US-0136021.

PR 28-MAY-1999; 99US-0136392.

PR 01-JUN-1999; 99US-0136782.

PR 03-JUN-1999; 99US-0137222.

PR 04-JUN-1999; 99US-0137528.

PR 07-JUN-1999; 99US-0137502.

PR 08-JUN-1999; 99US-0137724.

PR 10-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 14-JUN-1999; 99US-0138847.

PR 16-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 17-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 77.4%; Score 41; DB 21; Length 353;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AOPCPVLS 8
| | | | |
Db 154 AOPCPVLS 161

RESULT 8
AAU00449

ID AU00449 standard; Protein; 362 AA.
XX
AC AA000449;
XX
DT 19-JUN-2001 (first entry)
XX
DE Maize disease resistance enhancing protein ZmPtl1-1.
XX
KW Maize; plant disease resistance; crop; soybean; sunflower; sorghum;
KW canola; wheat; alfalfa; cotton; rice; barley; millet; plant cell death;
KW herbicide resistance; ZmPtl1-1; Ptl1; Ptol.
XX
OS Zea mays.
XX
PN WO200118061-A2.
XX
PD 15-MAR-2001.
XX
PF 06-SEP-2000; 2000WO-US24403.
XX
PR 09-SEP-1999; 99US-0152988.
XX
PA (P10N-) PIONEER HI-BRED INT INC.
XX
PI Simmons CR;
XX
DR WPI: 2001-226742/23.
DR N-PSDB: AAS01019.
XX
PT Novel isolated maize disease resistance polynucleotide useful for
PT increasing resistance in a plant to disease, controlling cell death,
PT and conferring resistance to herbicides -
XX
PS Claim 11; Page 89-90; 90pp; English.
XX
CC The present sequence representing maize ZmPtl1-1 protein is 1 of 7
CC novel disease resistance proteins (AAU00443-AAU00449). The ZmPtl1-1 is
CC a Zea mays gene homologue to the plant disease resistance genes Ptl1
CC and Ptol in tomato. These novel maize disease resistance polynucleotides
CC and polypeptides are useful for enhancing disease resistance in crops
CC and transgenic plants including maize, soybean, sunflower, sorghum,
CC canola, wheat, alfalfa, cotton, rice, barley or millet. The level of
CC disease resistance protein is increased by transforming a plant cell
CC with a recombinant expression cassette which comprises the disease
CC resistance polynucleotide operably linked to a promoter, or by culturing
CC the plant cell under plant growing conditions to produce a regenerated
CC plant, or by inducing expression of the polynucleotide to modulate the
CC disease resistance protein in a plant. The polynucleotides encoding the
CC disease resistance proteins are useful for increasing resistance in a
CC plant to disease, controlling cell death, and conferring resistance to
CC herbicides. They are useful as probes or amplification primers in the
CC detection, quantification, or isolation of gene transcripts. They can be
CC used for recombinant expression or isolation of gene transcripts, as
CC immunogens in the preparation and/or screening of antibodies, and in
CC sense or antisense suppression of the polynucleotide in a host cell,
CC tissue or plant.
XX
SQ Sequence 362 AA;
Query Match 77.4%; Score 41; DB 22; Length 362;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 35641.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP103405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.

RESULT 9
AAG29892
ID AAG29892 standard; Protein; 365 AA.
XX
AC AAG29892;

PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140685.
PR 24-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142330.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144864.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 25-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150864.

PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 77.4%; Score 41; DB 21; Length 365;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AOPCPVLS 8
Db 166 AOPCPVLS 173

RESULT 10
AAG29891
ID AAG29891 standard; Protein: 397 AA.

XX AAG29891;
AC
XX
DT 17-OCT-2000 (first entry)
XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 35640.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.
OS Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143564.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145921.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.

PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161982.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 77.4%; Score 41; DB 21; Length 397;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQP6PVLS 8
Db 198 AQP6PVLS 205

RESULT 11
AAW97359
ID AAW97359 standard; Protein; 592 AA.

AC AAW97359;

DT 12-MAY-1999 (first entry)

DE A 6-Transmembrane protein designated SDR2.

KW 6-Transmembrane protein; SDR2; cancer; inflammation; autoimmunity;
KW allergy; asthma; rheumatoid arthritis; CNS inflammation;
KW cerebellar degeneration; Alzheimer's disease; Parkinson's disease;
KW multiple sclerosis; amyotrophic lateral sclerosis; head injury damage;
KW neurological abnormality; septic shock; sepsis; stroke; osteoporosis;
KW osteoarthritis; ischaemia reperfusion injury; cardiovascular disease;
KW kidney disease; liver disease; ischaemic injury; myocardial infarction;
KW hypotension; hypertension; AIDS; myelodysplastic syndrome;
KW hematological abnormality; aplastic anaemia; male pattern baldness;

KW Infection.
XX
OS Homo sapiens.
XX
PN EP897979-A2.
XX
PD 24-FEB-1999.
XX
PF 06-AUG-1998; 98EP-0306292.
XX
PR 16-DEC-1997; 97US-0991813.
PR 12-AUG-1997; 97US-0055375.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI Albion EF, Kikly KK;
XX WPI; 1999-134642/12.
DR N-PSDB; MAX15876.
XX
PT New 6-Transmembrane protein (SDR2) polypeptide and polynucleotide --
PT useful as diagnostic reagents and for prevention and treatment of
PT cancer, AIDS and microbial infections
XX
PS Claim 2; Page 6; 19pp; English.

CC The present sequence represents a 6-Transmembrane protein polypeptide
CC designated SDR2. SDR2 polypeptides and polynucleotides are useful for
CC diagnosing susceptibility to diseases by detecting mutations in the SDR2
CC gene, and can diagnose diseases associated with SDR2 protein imbalance by
CC determining SDR2 polypeptide expression levels. SDR2 polypeptides can be
CC used to screen for agonists and antagonists and to produce antibodies.
CC Diseases diagnosed, prevented and treated include: cancer, inflammation,
CC autoimmunity, allergy, asthma, rheumatoid arthritis, CNS inflammation,
CC cerebellar degeneration, Alzheimer's disease, Parkinson's disease,
CC multiple sclerosis, amyotrophic lateral sclerosis, head injury damage,
CC and other neurological abnormalities, septic shock, sepsis, stroke,
CC osteoporosis, osteoarthritis, ischaemia reperfusion injury,
CC cardiovascular disease, kidney disease, liver disease, ischaemic injury,
CC myocardial infarction, hypotension, hypertension, AIDS, myelodysplastic
CC syndromes and other hematological abnormalities, aplastic anaemia, male
CC pattern baldness, and bacterial, fungal, protozoan and viral infections.
XX
SQ Sequence 592 AA;

Query Match 77.4%; Score 41; DB 20; Length 592;
Best Local Similarity 75.0%; Pred. No. 74;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 PGPVLSQP 10
Db 158 PGPVLSQP 165

RESULT 12
ABP11105
ID ABP11105 standard; Protein; 63 AA.

AC ABP11105;

DT 24-JUN-2002 (first entry)

DE Human ORF protein sequence SRQ ID NO:22192.

KW Human; open reading frame; ORF; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.

PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0234984.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241788.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246509.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.

PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-451937/48.
 XX N-PSDB; AAL35312.
 PT
 PT Isolated polypeptide for treating, preventing and/or prognosing
 PT disorders related to the musculoskeletal system including
 PT musculoskeletal cancers and also for testing and detection e.g.
 PT diagnosis -
 XX
 PS Claim 11; SEQ ID NO 1677; 781bp + Sequence Listing; English.
 CC
 CC The invention relates to novel genes (AAL34669-AAL37666) and proteins
 CC (ABB03087-ABB04109) associated with the musculoskeletal system useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (antagonists are useful in the diagnosis, treatment
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
 CC other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC (c) cardiovascular disorders such as myocardial ischemias; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 CC
 CC Sequence 113 AA;
 QQ
 Db 102 QPGPTLRQP 110
 QQ 2 QPGPTLRQP 10
 |||||
 Db 102 QPGPTLRQP 110
 Query Match 75.5%; Score 40; DB 22; Length 113;
 Best Local Similarity 77.8%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 RESULT 14
 ABG20476
 ID ABG20476 standard; Protein: 199 AA.
 XX
 AC ABG20476;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #20467.
 DE
 XX
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 OS Homo sapiens.
 OS
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX

PA (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX
 XX WPI: 2001-639362/73.
 DR N-PSDB: AAS84663.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20: SEQ ID No 50835; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG3037 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 199 AA:
 QY
 Db 173 AOPGPVLSQP 10
 173 AOPGPVGSNP 182
 RESULT 15
 AAE03611
 ID AAE03611 standard; peptide; 24 AA.
 XX
 AC AAE03611;
 XX
 DT 07-AUG-2001 (first entry)
 XX
 DE Human AZAD protein leucine-rich repeat sequence #7.
 XX
 KW Human; leucine-rich repeat; LRR; AZAD; neurodegenerative disorder;
 KW CNS disorder; central nervous system disorder; prostate disorder;
 KW prostatic; benign prostatic hyperplasia; adenocarcinoma;
 KW prostate cancer; genitourinary system carcinoma; testicular tumour;
 KW Alzheimer's disease; dementia; Parkinson's disease; multiple sclerosis;
 KW amyotrophic lateral sclerosis; epilepsy; psychiatric disorder; mania;
 KW depression; schizophrenia; anxiety; phobic disorder; learning disorder;
 KW memory disorder; amnesia; migraine; protein-protein interaction;
 KW cellular activity; neoplastic transformation; drug screening;
 KW forensic identification; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200142286-A2.
 XX
 PD 14-JUN-2001.

XX
 PF 07-DEC-2000; 2000WO-0533140.
 XX
 PR 08-DEC-1999; 99US-0456592.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Rhodadoust MM;
 XX
 DR WPI: 2001-381633/40.
 XX
 PT New leucine-rich repeat-containing protein and nucleic acid molecules
 PT for diagnosing, treating neural disorders, such as neurodegenerative
 PT disorders, such as Alzheimer's disease, dementia, epilepsy and prostate
 PT cancer -
 XX
 PS Example 1; Fig 4; 133pp; English.
 XX
 CC The present sequence is a leucine-rich repeat of
 CC human AZAD protein which is a multiple leucine-rich repeat-containing
 CC secreted protein. AZAD polypeptide is useful for
 CC identifying a compound which modulates its activity and binds to it.
 CC AZAD proteins and nucleic acid molecules are useful for treating and
 CC diagnosing AZAD-mediated or related disorders, which includes a neural
 CC disorder (e.g. neurodegenerative disorders, including CNS disorders)
 CC and a prostate disorder, e.g. prostaticitis, benign prostatic hyperplasia
 CC and cancer (e.g. adenocarcinoma, prostate cancer, genitourinary system
 CC carcinomas and testicular tumour). The neurodegenerative disorders
 CC include Alzheimer's disease, dementias related to Alzheimer's disease,
 CC Parkinson's disease, multiple sclerosis, amyotrophic lateral sclerosis,
 CC epilepsy, psychiatric disorders e.g. depression, schizophrenia
 CC disorders e.g. mania, anxiety or phobic disorders, learning or memory
 CC disorders e.g. amnesia or age-related memory loss, and neurological
 CC disorders e.g. migraine. AZAD molecules are useful as markers of
 CC disorders or disease states, as markers for precursors of disease state,
 CC for predisposition of disease state or as markers of drug activity or
 CC pharmacogenomic profile of a subject. The AZAD polypeptide and
 CC polynucleotide are capable of modulating protein-protein interaction,
 CC e.g. by interacting with an extracellular component, thereby modulating
 CC cellular activities, including attachment, adhesion, migration,
 CC patterning, growth and/or differentiation of a cell. AZAD proteins
 CC regulate embryonic development and differentiation, tissue maintenance
 CC and function, pathological conditions, e.g. neuronal degeneration,
 CC neoplastic transformation and tumour progression. AZAD proteins are
 CC useful as immunogens to raise anti-AZAD antibodies which are useful to
 CC detect and isolate AZAD proteins and modulate its activity. AZAD
 CC proteins are useful to screen for naturally occurring AZAD substrates
 CC and to screen for drugs or compounds which modulate AZAD activity.
 CC AZAD nucleic acid fragments are useful as primers or hybridisation
 CC probes for the detection of AZAD-encoding nucleic acids. cDNA encoding
 CC the AZAD protein are useful in gene therapy. AZAD sequences are also
 CC useful to map their respective genes on a chromosome, for tissue typing
 CC and in forensic identification of a biological sample.
 XX
 SQ Sequence 24 AA:
 QY
 Db 14 ALRPPVLSQ 22
 1 AOPGPVLSQ 9
 14 ALRPPVLSQ 22
 Query Match 73.6%; Score 39; DB 22; Length 24;
 Best Local Similarity 88.9%; Pred. No. 7;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Search completed: February 16, 2003, 21:55:08
 Job time : 2.7666 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:59:29 ; Search time 0.256959 Seconds
(without alignments)
1145.042 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_452
Perfect score: 53
Sequence: 1 AOPGVLSP 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/55_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/55_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/55_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/55_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/55_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/55_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	100.0	801	1	US-08-725-012-2
2	41	77.4	592	3	US-08-991-813-2
3	38	71.7	942	3	US-09-074-579-1
4	38	71.7	942	4	US-09-388-774-1
5	37	69.8	186	4	US-09-325-932A-72
6	37	69.8	349	4	US-09-161-241-8
7	37	69.8	645	4	US-08-791-115B-6
8	37	69.8	756	2	US-08-887-518-4
9	37	69.8	756	2	US-09-023-321-4
10	37	69.8	756	2	US-08-890-853-2
11	37	69.8	756	2	US-09-032-475-4
12	37	69.8	756	2	US-09-099-125A-2
13	37	69.8	756	2	US-09-099-124A-2
14	37	69.8	756	4	US-09-032-476-2
15	37	69.8	756	4	US-08-890-854-2
16	37	69.8	756	4	US-09-023-324-2
17	37	69.8	756	4	US-09-168-629-15
18	37	69.8	756	4	US-08-910-820-9
19	36	67.9	362	2	US-09-055-097-7
20	36	67.9	404	3	US-08-459-953A-7
21	35	66.0	155	4	US-08-630-915A-10
22	35	66.0	350	4	US-09-161-241-9
23	34	64.2	249	4	US-09-345-468-9
24	34	64.2	249	4	US-09-414-453A-9
25	34	64.2	307	4	US-09-134-001C-3238
26	34	64.2	319	4	US-09-345-468-5
27	34	64.2	319	4	US-09-414-453A-5

28	34	64.2	328	2	US-08-651-818A-2	Sequence 2, Appl1
29	34	64.2	328	4	US-09-184-826-2	Sequence 2, Appl1
30	34	64.2	339	4	US-09-345-468-3	Sequence 3, Appl1
31	34	64.2	339	4	US-09-414-453A-3	Sequence 3, Appl1
32	34	64.2	384	1	US-08-221-750A-11	Sequence 11, Appl1
33	34	64.2	565	4	US-08-906-156A-12	Sequence 12, Appl1
34	34	64.2	742	4	US-08-791-115B-5	Sequence 5, Appl1
35	34	64.2	830	4	US-09-562-737-38	Sequence 38, Appl1
36	34	64.2	2627	2	US-08-751-189-3	Sequence 3, Appl1
37	34	64.2	2627	2	US-09-060-836-3	Sequence 3, Appl1
38	34	64.2	2627	3	US-09-184-445-3	Sequence 3, Appl1
39	33	62.3	222	3	US-08-476-509B-40	Sequence 40, Appl1
40	33	62.3	176	2	US-08-619-542B-44	Sequence 44, Appl1
41	33	62.3	180	3	US-09-133-341-12	Sequence 12, Appl1
42	33	62.3	444	4	US-09-129-668-8	Sequence 8, Appl1
43	33	62.3	464	2	US-08-969-630-5	Sequence 5, Appl1
44	33	62.3	580	3	US-08-906-865-1	Sequence 1, Appl1
45	33	62.3	580	4	US-09-129-668-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-08-725-012-2
Sequence 2, Application US/08725012
Patent No. 5710025
GENERAL INFORMATION:
APPLICANT: Dikstein, Rivka
APPLICANT: Tian, Robert
TITLE OF INVENTION: B-Cell Specific Transcription Factor
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/725,012
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: B97-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 801 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-725-012-2
Query Match 100.0%; Score 53; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AOPGVLSP 10
DB 443 AOPGVLSP 452
RESULT 2

US-08-991-813-2
; Sequence 2, Application US/08991813
; Patent No. 6050579
; GENERAL INFORMATION:
; APPLICANT: ALBONE, EARL
; APPLICANT: KIKLY, KRISTINE
; TITLE OF INVENTION: HUMAN SDR2 CDNA CLONE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,813
; FILING DATE: 16-DEC-1997
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 60/055,375
; FILING DATE: 12-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH-70281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 592 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-991-813-2

Query Match 77.4%; Score 41; DB 3; Length 592;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 QGPVLSQP 10
|||||
Db 158 QGPVLSQP 165

RESULT 3
US-09-074-579-1
; Sequence 1, Application US/09074579
; Patent No. 6001596
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE
; TITLE OF INVENTION: INHIBITOR HEAVY CHAIN PRECURSOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Word Perfect 6.1/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/074,579
FILING DATE: HEREWITH
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Cerrone, Michael C
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0505 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 942 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: UTRSN0702
CLONE: 688183

US-09-074-579-1

Query Match 71.7%; Score 38; DB 3; Length 942;
Best Local Similarity 66.7%; Pred. No. 11e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 QGPVLSQP 10
|||||
Db 658 QGPVLSQP 666

RESULT 4
US-09-388-774-1
; Sequence 1, Application US/09388774
; Patent No. 6228991
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE
; TITLE OF INVENTION: INHIBITOR HEAVY CHAIN PRECURSOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Word Perfect 6.1/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/388,774
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 09/074,579
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Cerrone, Michael C
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0505 US
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 942 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: UTRSN0702
CLONE: 688183
US-09-388-774-1

Query Match 71.7%; Score 38; DB 4; Length 942;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 QPGVLSOP 10
Db 658 QPGPLKKP 666

RESULT 5
US-09-325-932A-72
Sequence 72, Application US/09325932A
Patent No. 6451604
GENERAL INFORMATION:
APPLICANT: Flihn, Barry
APPLICANT: Lasham, Annette
TITLE OF INVENTION: Compositions affecting programmed cell
FILE REFERENCE: 1022
CURRENT APPLICATION NUMBER: US/09/325,932A
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 206
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 72
LENGTH: 186
TYPE: PRT
ORGANISM: Pinus radiata
US-09-325-932A-72

Query Match 69.8%; Score 37; DB 4; Length 186;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AQPGLVL 7
Db 166 AQPGLVL 172

RESULT 6
US-09-161-241-8
Sequence 8, Application US/09161241
Patent No. 6344541
GENERAL INFORMATION:
APPLICANT: Bass, Michael B
APPLICANT: Sullivan, John K
APPLICANT: Theill, Lars E
APPLICANT: Wang, Daquan
TITLE OF INVENTION: NOVEL DKR POLYPEPTIDES
FILE REFERENCE: A-548
CURRENT APPLICATION NUMBER: US/09/161,241
CURRENT FILING DATE: 1998-09-25
NUMBER OF SEQ ID NOS: 78
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 8
LENGTH: 349
TYPE: PRT
ORGANISM: Mouse
US-09-161-241-8

Query Match 69.8%; Score 37; DB 4; Length 349;
Best Local Similarity 60.0%; Pred. No. 56;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 AQPGLVL 10
Db 32 AEPGLALNP 41

RESULT 7
US-08-791-115B-6
Sequence 6, Application US/08791115B
Patent No. 6262242
GENERAL INFORMATION:
APPLICANT: Steck, Peter
APPLICANT: Pershouse, Mark A.
APPLICANT: Jasser, Samir
APPLICANT: Yung, W.K. Alfred
TITLE OF INVENTION: A TUMOR SUPPRESSOR DESIGNATED TS10Q23.3
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
STREET: 555 Thirteenth Street, N.W., Suite 701-E
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 22204
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/791,115B
FILING DATE: 30-JAN-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 38,957
REFERENCE/DOCKET NUMBER: 2318-134.A
TELEPHONE: 202-683-6040
TELEFAX: 202-683-7031
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 645 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-791-115B-6

Query Match 69.8%; Score 37; DB 4; Length 645;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 PGPVLSOP 10
Db 126 PGPVLSLP 133

RESULT 8
US-08-887-518-4
Sequence 4, Application US/08887518
Patent No. 5843721
GENERAL INFORMATION:
APPLICANT: Rothe, Mike
APPLICANT: Wu, Lin
TITLE OF INVENTION: NIK Proteins, Nucleic Acids and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200

CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/887,518
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: T97-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ. ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 756 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-887-518-4

Query Match
Best Local Similarity 69.8%; Score 37; DB 2; Length 756;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 AOPGPVLSQP 10
:||||:||||
DB 682 SGPGLMSQP 691

RESULT 9
US-09-023-321-4
Sequence 4, Application US/09023321
Patent No. 5844073
GENERAL INFORMATION:
APPLICANT: Rolhe, Mike
TITLE OF INVENTION: NIK Proteins, Nucleic Acids and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,321
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/887,518
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: T97-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342

INFORMATION FOR SEQ. ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 756 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-023-321-4

Query Match
Best Local Similarity 69.8%; Score 37; DB 2; Length 756;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 AOPGPVLSQP 10
:||||:||||
DB 682 SGPGLMSQP 691

RESULT 10
US-08-890-853-2
Sequence 2, Application US/08890853
Patent No. 5851812
GENERAL INFORMATION:
APPLICANT: Goeddel, David V.
TITLE OF INVENTION: IKK- Proteins, Nucleic Acids and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/890,853
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: T97-006-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 756 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-890-853-2

Query Match
Best Local Similarity 69.8%; Score 37; DB 2; Length 756;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 AOPGPVLSQP 10
:||||:||||
DB 682 SGPGLMSQP 691

RESULT 11
US-09-032-475-4
Sequence 4, Application US/09032475
Patent No. 5854003
GENERAL INFORMATION:

APPLICANT: Rothe, Mike
TITLE OF INVENTION: NIK Proteins, Nucleic Acids and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,475
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/887,518
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: T97-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 756 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-032-475-4
Query Match 69.8%; Score 37; DB 2; Length 756;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 AQPGLVLSOP 10
Db 682 SQPGQLMSOP 691
RESULT 12
US-09-099-125A-2
Sequence 2, Application US/09099125A
Patent No. 5916760
GENERAL INFORMATION:
APPLICANT: Goeddel, David V.
APPLICANT: Moronitz, John
TITLE OF INVENTION: IKK- Proteins, Nucleic Acids and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/099,125A
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/890,853
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: T97-006-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 756 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-099-125A-2
Query Match 69.8%; Score 37; DB 2; Length 756;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 AQPGLVLSOP 10
Db 682 SQPGQLMSOP 691
RESULT 13
US-09-099-124A-2
Sequence 2, Application US/09099124A
Patent No. 5939302
GENERAL INFORMATION:
APPLICANT: Goeddel, David V.
APPLICANT: Moronitz, John
TITLE OF INVENTION: IKK- Proteins, Nucleic Acids and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/099,124A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/890,853
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: T97-006-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 756 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-099-124A-2
Query Match 69.8%; Score 37; DB 2; Length 756;

Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 AOPGPVLSOP 10
:||||:||||
DB 682 SOPGQLMSOP 691

RESULT 14

US-09-032-476-2
; Sequence 2, Application US/09032476
; Patent No. 6235492
; GENERAL INFORMATION:
; APPLICANT: Rothe, Mike
; APPLICANT: Cao, Zhaoan
; APPLICANT: R guler, Catherine
; TITLE OF INVENTION: IKK- Proteins, Nucleic Acids and Methods
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,476
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/890,854
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: T97-006-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 756 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-032-476-2

Query Match 69.8%; Score 37; DB 4; Length 756;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 AOPGPVLSOP 10
:||||:||||
DB 682 SOPGQLMSOP 691

RESULT 15

US-08-890-854-2
; Sequence 2, Application US/08890854
; Patent No. 6235512
; GENERAL INFORMATION:
; APPLICANT: Rothe, Mike
; APPLICANT: Cao, Zhaoan
; APPLICANT: R guler, Catherine
; TITLE OF INVENTION: IKK- Proteins, Nucleic Acids and Methods
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:

ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/890,854
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: T97-006-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 756 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-890-854-2

Query Match 69.8%; Score 37; DB 4; Length 756;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 AOPGPVLSOP 10
:||||:||||
DB 682 SOPGQLMSOP 691

Search completed: February 16, 2003, 22:02:59
Job time: 2.25696 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: February 16, 2003, 21:56:34 ; Search time 0.197002 Seconds
(without alignments)
1296.883 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_452

Perfect score: 53

Sequence: 1 AQPGLVSLQP 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapect 0.5

Searched: 140259 seqs, 25548876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	75.5	113	10	US-09-764-877-1677
2	39	73.6	24	10	US-09-789-404-14
3	39	73.6	794	10	US-09-789-404-2
4	38	71.7	694	9	US-09-944-413-55
5	38	71.7	694	9	US-09-944-403-55
6	38	71.7	694	9	US-09-944-896-55
7	38	71.7	694	9	US-09-944-944-55
8	38	71.7	694	9	US-09-944-929-55
9	38	71.7	694	9	US-09-944-929-55
10	38	71.7	694	9	US-10-174-590-60
11	38	71.7	694	9	US-10-174-590-60
12	38	71.7	694	9	US-10-175-737-60
13	38	71.7	694	9	US-10-175-737-60
14	38	71.7	694	9	US-10-175-738-60
15	38	71.7	694	9	US-10-175-752-60
16	38	71.7	694	9	US-10-176-482-60
17	38	71.7	694	9	US-10-176-482-60
18	38	71.7	694	9	US-10-176-757-60
19	38	71.7	694	9	US-10-176-913-60

20	38	71.7	694	9	US-10-180-557-60	Sequence 60, App1
21	38	71.7	694	9	US-10-173-700-60	Sequence 60, App1
22	38	71.7	694	9	US-10-174-572-60	Sequence 60, App1
23	38	71.7	694	9	US-10-174-579-60	Sequence 60, App1
24	38	71.7	694	9	US-10-174-582-60	Sequence 60, App1
25	38	71.7	694	9	US-10-174-588-60	Sequence 60, App1
26	38	71.7	694	9	US-10-175-739-60	Sequence 60, App1
27	38	71.7	694	9	US-10-175-740-60	Sequence 60, App1
28	38	71.7	694	9	US-10-175-743-60	Sequence 60, App1
29	38	71.7	694	9	US-10-176-488-60	Sequence 60, App1
30	38	71.7	694	9	US-10-176-492-60	Sequence 60, App1
31	38	71.7	694	9	US-10-176-747-60	Sequence 60, App1
32	38	71.7	694	9	US-10-176-750-60	Sequence 60, App1
33	38	71.7	694	9	US-10-176-985-60	Sequence 60, App1
34	38	71.7	694	9	US-10-176-987-60	Sequence 60, App1
35	38	71.7	694	9	US-10-176-991-60	Sequence 60, App1
36	38	71.7	694	9	US-10-176-992-60	Sequence 60, App1
37	38	71.7	694	9	US-10-176-993-60	Sequence 60, App1
38	38	71.7	694	9	US-10-184-658-60	Sequence 60, App1
39	38	71.7	694	10	US-09-866-028-55	Sequence 55, App1
40	38	71.7	694	10	US-09-944-449-55	Sequence 55, App1
41	38	71.7	694	10	US-09-944-457-55	Sequence 55, App1
42	38	71.7	694	10	US-09-944-862-55	Sequence 55, App1
43	38	71.7	694	10	US-09-945-587-55	Sequence 55, App1
44	38	71.7	694	10	US-09-945-015-55	Sequence 55, App1
45	38	71.7	694	10	US-09-944-396-55	Sequence 55, App1

ALIGNMENTS

RESULT 1
US-09-764-877-1677
Sequence 1677, Application US/09764877
Patent No. US20020147140A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
FILE REFERENCE: PC005
CURRENT APPLICATION NUMBER: US/09/764,877
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 4031
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1677
LENGTH: 113
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (67)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (73)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-877-1677

Query Match 75.5% Score 40; DB 10; Length 113;
Best Local Similarity 77.8% Pred. No. 5.6;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AQPGLVSLQP 10
|||||
Db 102 QPGLTIRP 110

RESULT 2
US-09-789-404-14
Sequence 14, Application US/09789404
Patent No. US20020025554A1
GENERAL INFORMATION:
APPLICANT: Khodadoust, Mehran
TITLE OF INVENTION: NOVEL LEUCINE RICH REPEAT-CONTAINING MOLECULES AND USES THEREOF

FILE REFERENCE: 10448/008001
CURRENT APPLICATION NUMBER: US/09/789,404
CURRENT FILING DATE: 2001-02-20
PRIOR APPLICATION NUMBER: 09/456,592
PRIOR FILING DATE: 1999-12-08
NUMBER OF SEQ ID NOS: 30
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 24
TYPE: PRT
ORGANISM: Homo sapiens
US-09-789-404-14

Query Match
Best Local Similarity 73.6%; Score 39; DB 10; Length 24;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACPGVLISQ 9
DB 14 ALPGPVLSQ 22

RESULT 3
US-09-789-404-2
Sequence 2: Application US/09789404
Patent No. US2002002554A1
GENERAL INFORMATION:
APPLICANT: Knodasoust, Mehnan
TITLE OF INVENTION: NOVEL LEUDINE RICH REPEAT-CONTAINING MOLECULES AND USES THEREFOR
FILE REFERENCE: 10448/008001
CURRENT APPLICATION NUMBER: US/09/789,404
CURRENT FILING DATE: 2001-02-20
PRIOR APPLICATION NUMBER: 09/456,592
PRIOR FILING DATE: 1999-12-08
NUMBER OF SEQ ID NOS: 30
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 794
TYPE: PRT
ORGANISM: Homo sapiens
US-09-789-404-2

Query Match
Best Local Similarity 73.6%; Score 39; DB 10; Length 794;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACPGVLISQ 9
DB 253 ALPGPVLSQ 261

RESULT 4
US-09-944-413-55
Sequence 55: Application US/09944413
Patent No. US20020156004A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Botstein, David
APPLICANT: Baton, Dan
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerltsen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul
APPLICANT: Grimaldi, Christopher
APPLICANT: Gurney, Austin
APPLICANT: Hillan, Kenneth
APPLICANT: Kljavin, Ivar
APPLICANT: Napier, Mary
APPLICANT: Roy, Margaret
APPLICANT: Tumas, Daniel
APPLICANT: Wood, William
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P2548P1C1
CURRENT APPLICATION NUMBER: US/09/944,413
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 09/866,028
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: 60/067,411
PRIOR FILING DATE: December 3, 1997
PRIOR APPLICATION NUMBER: 60/069,334
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,335
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,278
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,425
PRIOR FILING DATE: December 12, 1997
PRIOR APPLICATION NUMBER: 60/069,696
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,694
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,702
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,870
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/069,873
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/068,017
PRIOR FILING DATE: December 18, 1997
PRIOR APPLICATION NUMBER: 60/070,440
PRIOR FILING DATE: January 5, 1998
PRIOR APPLICATION NUMBER: 60/074,086
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/074,092
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/075,945
PRIOR FILING DATE: February 25, 1998
PRIOR APPLICATION NUMBER: 60/112,850
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 60/113,296
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 60/146,222
PRIOR FILING DATE: July 28, 1999
PRIOR APPLICATION NUMBER: PCT/US98/19330
PRIOR FILING DATE: September 16, 1998
PRIOR APPLICATION NUMBER: PCT/US98/25108
PRIOR FILING DATE: December 1, 1998
PRIOR APPLICATION NUMBER: 09/216,021
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 09/218,517
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 09/254,311
PRIOR FILING DATE: March 3, 1999
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: June 22, 1999
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: September 15, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28409
PRIOR FILING DATE: NO. US20020156004A1
PRIOR FILING DATE: NO. US20020156004A1
PRIOR FILING DATE: NO. US20020156004A1
PRIOR FILING DATE: PCT/US99/28301
PRIOR FILING DATE: December 11, 1999
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: December 16, 1999
PRIOR APPLICATION NUMBER: PCT/US00/03565
PRIOR FILING DATE: February 11, 2000
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: February 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: March 2, 2000
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: March 30, 2000
PRIOR APPLICATION NUMBER: PCT/US00/14042

;; PRIOR FILING DATE: May 22, 2000
;; PRIOR APPLICATION NUMBER: PCT/US00/20710
;; PRIOR FILING DATE: July 28, 2000
;; PRIOR APPLICATION NUMBER: PCT/US00/32678
;; PRIOR FILING DATE: December 1, 2000
;; PRIOR APPLICATION NUMBER: PCT/US01/06520
;; PRIOR FILING DATE: February 28, 2001
;; NUMBER OF SEQ ID NOS: 120
;; SEQ ID NO 55
;; LENGTH: 694
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-09-944-413-55

Query Match 71.7%; Score 38; DB 9; Length 694;
Best Local Similarity 66.7%; Pred. No. 75;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 2 QPGVLSOP 10
Db 658 QPGPLKKP 666

RESULT 5
US-09-944-403-55
;; Sequence 55, Application US/09944403
;; Patent No. US20020165143A1
;; GENERAL INFORMATION:
;; APPLICANT: Baker, Kevin
;; APPLICANT: Botstein, David
;; APPLICANT: Eaton, Dan
;; APPLICANT: Ferrara, Napoleone
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Gerltsen, Mary
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Grimaldi, Christopher
;; APPLICANT: Gurney, Austin
;; APPLICANT: Hillan, Kenneth
;; APPLICANT: Kijavlin, Ivar
;; APPLICANT: Napier, Mary
;; APPLICANT: Roy, Margaret
;; APPLICANT: Tumas, Daniel
;; APPLICANT: Wood, William
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P2548P1C1
;; CURRENT APPLICATION NUMBER: US/09/944,403
;; PRIOR FILING DATE: 2001-09-26
;; PRIOR APPLICATION NUMBER: 09/866,028
;; PRIOR FILING DATE: 2001-05-25
;; PRIOR APPLICATION NUMBER: 60/067,411
;; PRIOR FILING DATE: December 3, 1997
;; PRIOR APPLICATION NUMBER: 60/069,334
;; PRIOR FILING DATE: December 11, 1997
;; PRIOR APPLICATION NUMBER: 60/069,335
;; PRIOR FILING DATE: December 11, 1997
;; PRIOR APPLICATION NUMBER: 60/069,278
;; PRIOR FILING DATE: December 11, 1997
;; PRIOR APPLICATION NUMBER: 60/069,425
;; PRIOR FILING DATE: December 12, 1997
;; PRIOR APPLICATION NUMBER: 60/069,696
;; PRIOR FILING DATE: December 16, 1997
;; PRIOR APPLICATION NUMBER: 60/069,694
;; PRIOR FILING DATE: December 16, 1997
;; PRIOR APPLICATION NUMBER: 60/069,702
;; PRIOR FILING DATE: December 16, 1997
;; PRIOR APPLICATION NUMBER: 60/069,870
;; PRIOR FILING DATE: December 17, 1997
;; PRIOR APPLICATION NUMBER: 60/069,873
;; PRIOR FILING DATE: December 17, 1997
;; PRIOR APPLICATION NUMBER: 60/068,017
;; PRIOR FILING DATE: December 18, 1997

;; PRIOR APPLICATION NUMBER: 60/070,440
;; PRIOR FILING DATE: January 5, 1998
;; PRIOR APPLICATION NUMBER: 60/074,086
;; PRIOR FILING DATE: February 9, 1998
;; PRIOR APPLICATION NUMBER: 60/074,092
;; PRIOR FILING DATE: February 9, 1998
;; PRIOR APPLICATION NUMBER: 60/075,945
;; PRIOR FILING DATE: February 25, 1998
;; PRIOR APPLICATION NUMBER: 60/112,850
;; PRIOR FILING DATE: December 16, 1998
;; PRIOR APPLICATION NUMBER: 60/113,296
;; PRIOR FILING DATE: December 22, 1998
;; PRIOR APPLICATION NUMBER: 60/146,222
;; PRIOR FILING DATE: July 28, 1999
;; PRIOR APPLICATION NUMBER: PCT/US98/19330
;; PRIOR FILING DATE: September 16, 1998
;; PRIOR APPLICATION NUMBER: PCT/US98/25108
;; PRIOR FILING DATE: December 1, 1998
;; PRIOR APPLICATION NUMBER: 09/216,021
;; PRIOR FILING DATE: December 16, 1998
;; PRIOR APPLICATION NUMBER: 09/218,517
;; PRIOR FILING DATE: December 22, 1998
;; PRIOR APPLICATION NUMBER: 09/254,311
;; PRIOR FILING DATE: March 3, 1999
;; PRIOR APPLICATION NUMBER: PCT/US99/12252
;; PRIOR FILING DATE: June 22, 1999
;; PRIOR APPLICATION NUMBER: PCT/US99/21090
;; PRIOR FILING DATE: September 15, 1999
;; PRIOR APPLICATION NUMBER: PCT/US99/28409
;; PRIOR FILING DATE: NO. US20020165143A1ember 30, 1999
;; PRIOR APPLICATION NUMBER: PCT/US99/28313
;; PRIOR FILING DATE: NO. US20020165143A1ember 30, 1999
;; PRIOR APPLICATION NUMBER: PCT/US99/28301
;; PRIOR FILING DATE: December 1, 1999
;; PRIOR APPLICATION NUMBER: PCT/US99/30095
;; PRIOR FILING DATE: December 16, 1999
;; PRIOR APPLICATION NUMBER: PCT/US00/03565
;; PRIOR FILING DATE: February 11, 2000
;; PRIOR APPLICATION NUMBER: PCT/US00/04414
;; PRIOR FILING DATE: February 22, 2000
;; PRIOR APPLICATION NUMBER: PCT/US00/05841
;; PRIOR FILING DATE: March 2, 2000
;; PRIOR APPLICATION NUMBER: PCT/US00/08439
;; PRIOR FILING DATE: March 30, 2000
;; PRIOR APPLICATION NUMBER: PCT/US00/14042
;; PRIOR FILING DATE: May 22, 2000
;; PRIOR APPLICATION NUMBER: PCT/US00/20710
;; PRIOR FILING DATE: July 28, 2000
;; PRIOR APPLICATION NUMBER: PCT/US00/32678
;; PRIOR FILING DATE: December 1, 2000
;; PRIOR APPLICATION NUMBER: PCT/US01/06520
;; PRIOR FILING DATE: February 28, 2001
;; NUMBER OF SEQ ID NOS: 120
;; SEQ ID NO 55
;; LENGTH: 694
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-09-944-403-55

Query Match 71.7%; Score 38; DB 9; Length 694;
Best Local Similarity 66.7%; Pred. No. 75;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 2 QPGVLSOP 10
Db 658 QPGPLKKP 666

RESULT 6
US-09-944-896-55
;; Sequence 55, Application US/09944896
;; Patent No. US20020168715A1
;; GENERAL INFORMATION:

APPLICANT: Baker, Kevin
APPLICANT: Botstein, David
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerlitsen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, Christopher
APPLICANT: Gurney, Austin
APPLICANT: Hillan, Kenneth
APPLICANT: Kijavlin, Ivar
APPLICANT: Napier, Mary
APPLICANT: Roy, Margaret
APPLICANT: Tumas, Daniel
APPLICANT: Wood, William
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P2548PICI
CURRENT APPLICATION NUMBER: US/09/944,896
CURRENT FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: 09/866,028
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: 60/069,334
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,335
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,278
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,425
PRIOR FILING DATE: December 12, 1997
PRIOR APPLICATION NUMBER: 60/069,696
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,694
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,702
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,870
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/069,873
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/068,017
PRIOR FILING DATE: December 18, 1997
PRIOR APPLICATION NUMBER: 60/070,440
PRIOR FILING DATE: January 5, 1998
PRIOR APPLICATION NUMBER: 60/074,086
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/074,092
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/075,945
PRIOR FILING DATE: February 25, 1998
PRIOR APPLICATION NUMBER: 60/112,850
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 60/113,296
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 60/146,222
PRIOR FILING DATE: July 28, 1999
PRIOR APPLICATION NUMBER: PCT/US98/19330
PRIOR FILING DATE: September 16, 1998
PRIOR APPLICATION NUMBER: PCT/US98/25108
PRIOR FILING DATE: December 1, 1998
PRIOR APPLICATION NUMBER: 09/216,021
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 09/218,517
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 09/254,311
PRIOR FILING DATE: March 3, 1999
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: June 22, 1999
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: September 15, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28409
PRIOR FILING DATE: NO. US20020168715A, member 30, 1999

PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: NO. US20020168715A, member 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28301
PRIOR FILING DATE: December 1, 1999
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: December 16, 1999
PRIOR APPLICATION NUMBER: PCT/US00/03565
PRIOR FILING DATE: February 11, 2000
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: February 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: March 2, 2000
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: March 30, 2000
PRIOR APPLICATION NUMBER: PCT/US00/14042
PRIOR FILING DATE: May 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/20710
PRIOR FILING DATE: July 28, 2000
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: December 1, 2000
PRIOR APPLICATION NUMBER: PCT/US01/06520
PRIOR FILING DATE: February 28, 2001
NUMBER OF SEQ ID NOS: 120
SEQ ID NO 55
LENGTH: 694
TYPE: PRT
ORGANISM: Homo Saplen
US-09-944-896-55
Query Match 71.7%; Score 38; DB 9; Length 694;
Best Local Similarity 66.7%; Pred. No. 75;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 QPGVLSQP 10
db 658 QPGVLSKP 666
RESULT 7
US-09-944-944-55
Sequence 55, Application US/09944944
Patent No. US20020173463A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Botstein, David
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerlitsen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, Christopher
APPLICANT: Gurney, Austin
APPLICANT: Hillan, Kenneth
APPLICANT: Kijavlin, Ivar
APPLICANT: Napier, Mary
APPLICANT: Roy, Margaret
APPLICANT: Tumas, Daniel
APPLICANT: Wood, William
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P2548PICI
CURRENT APPLICATION NUMBER: US/09/944,944
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 09/866,028
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: 60/067,411
PRIOR FILING DATE: December 3, 1997
PRIOR APPLICATION NUMBER: 60/069,334
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,335
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,278

;; PRIOR FILING DATE: December 11, 1997
;; PRIOR APPLICATION NUMBER: 60/069,425
;; PRIOR FILING DATE: December 12, 1997
;; PRIOR APPLICATION NUMBER: 60/069,696
;; PRIOR FILING DATE: December 16, 1997
;; PRIOR APPLICATION NUMBER: 60/069,694
;; PRIOR FILING DATE: December 16, 1997
;; PRIOR APPLICATION NUMBER: 60/069,702
;; PRIOR FILING DATE: December 16, 1997
;; PRIOR APPLICATION NUMBER: 60/069,870
;; PRIOR FILING DATE: December 17, 1997
;; PRIOR APPLICATION NUMBER: 60/069,873
;; PRIOR FILING DATE: December 17, 1997
;; PRIOR APPLICATION NUMBER: 60/068,017
;; PRIOR FILING DATE: December 18, 1997
;; PRIOR APPLICATION NUMBER: 60/070,440
;; PRIOR FILING DATE: January 5, 1998
;; PRIOR APPLICATION NUMBER: 60/074,086
;; PRIOR FILING DATE: February 9, 1998
;; PRIOR APPLICATION NUMBER: 60/074,092
;; PRIOR FILING DATE: February 9, 1998
;; PRIOR APPLICATION NUMBER: 60/075,945
;; PRIOR FILING DATE: February 25, 1998
;; PRIOR APPLICATION NUMBER: 60/112,850
;; PRIOR FILING DATE: December 16, 1998
;; PRIOR APPLICATION NUMBER: 60/113,296
;; PRIOR FILING DATE: December 22, 1998
;; PRIOR APPLICATION NUMBER: 60/146,222
;; PRIOR FILING DATE: July 28, 1999
;; PRIOR APPLICATION NUMBER: PCT/US98/19330
;; PRIOR FILING DATE: September 16, 1998
;; PRIOR APPLICATION NUMBER: PCT/US98/25108
;; PRIOR FILING DATE: December 1, 1998
;; PRIOR APPLICATION NUMBER: 09/216,021
;; PRIOR FILING DATE: December 16, 1998
;; PRIOR APPLICATION NUMBER: 09/218,517
;; PRIOR FILING DATE: December 22, 1998
;; PRIOR APPLICATION NUMBER: 09/254,311
;; PRIOR FILING DATE: March 3, 1999
;; PRIOR APPLICATION NUMBER: PCT/US99/12252
;; PRIOR FILING DATE: June 22, 1999
;; PRIOR APPLICATION NUMBER: PCT/US99/21090
;; PRIOR FILING DATE: September 15, 1999
;; PRIOR APPLICATION NUMBER: PCT/US99/28409
;; PRIOR FILING DATE: NO. US200201734631alembor 30, 1999
;; PRIOR APPLICATION NUMBER: PCT/US99/28313
;; PRIOR FILING DATE: NO. US200201734631alembor 30, 1999
;; PRIOR APPLICATION NUMBER: PCT/US99/28301
;; PRIOR FILING DATE: December 1, 1999
;; PRIOR APPLICATION NUMBER: PCT/US99/30095
;; PRIOR FILING DATE: December 16, 1999
;; PRIOR APPLICATION NUMBER: PCT/US00/03565
;; PRIOR FILING DATE: February 11, 2000
;; PRIOR APPLICATION NUMBER: PCT/US00/04414
;; PRIOR FILING DATE: February 22, 2000
;; PRIOR APPLICATION NUMBER: PCT/US00/05841
;; PRIOR FILING DATE: March 2, 2000
;; PRIOR APPLICATION NUMBER: PCT/US00/08439
;; PRIOR FILING DATE: March 30, 2000
;; PRIOR APPLICATION NUMBER: PCT/US00/14042
;; PRIOR FILING DATE: May 22, 2000
;; PRIOR APPLICATION NUMBER: PCT/US00/20710
;; PRIOR FILING DATE: July 28, 2000
;; PRIOR APPLICATION NUMBER: PCT/US00/32678
;; PRIOR FILING DATE: December 1, 2000
;; PRIOR APPLICATION NUMBER: PCT/US01/06520
;; PRIOR FILING DATE: February 28, 2001
;; NUMBER OF SEQ ID NOS: 120
;; SEQ ID NO 55
;; LENGTH: 694
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-09-944-944-55

Query Match 71.7%; Score 38; DB 9; Length 694;
Best Local Similarity 66.7%; Pred. No. 75;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 QPGPVLSP 10
||||:|
Db 658 QPGPLKKP 666

RESULT 8

US-09-944-907-55

;; Sequence 55, Application US/09944907
;; Publication No. US20020198147A1

GENERAL INFORMATION:

;; APPLICANT: Baker, Kevin
;; APPLICANT: Botstein, David
;; APPLICANT: Eaton, Dan
;; APPLICANT: Ferrara, Napoleone
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Gerritsen, Mary
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul
;; APPLICANT: Grimaldi, Christopher
;; APPLICANT: Gurney, Austin
;; APPLICANT: Hillan, Kenneth
;; APPLICANT: Kijavlin, Ivar
;; APPLICANT: Napier, Mary
;; APPLICANT: Roy, Margaret
;; APPLICANT: Tumas, Daniel
;; APPLICANT: Wood, William
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P2548P1C1
;; CURRENT APPLICATION NUMBER: US/09/944,907
;; CURRENT FILING DATE: 2001-08-31
;; PRIOR APPLICATION NUMBER: 09/866,028
;; PRIOR FILING DATE: 2001-05-25
;; NUMBER OF SEQ ID NOS: 120
;; SEQ ID NO 55
;; LENGTH: 694
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-09-944-907-55

Query Match 71.7%; Score 38; DB 9; Length 694;
Best Local Similarity 66.7%; Pred. No. 75;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 QPGPVLSP 10
||||:|
Db 658 QPGPLKKP 666

RESULT 9

US-09-944-929-55

;; Sequence 55, Application US/09944929
;; Publication No. US20020197612A1

GENERAL INFORMATION:

;; APPLICANT: Baker, Kevin
;; APPLICANT: Botstein, David
;; APPLICANT: Eaton, Dan
;; APPLICANT: Ferrara, Napoleone
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Gerritsen, Mary
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul
;; APPLICANT: Grimaldi, Christopher
;; APPLICANT: Gurney, Austin
;; APPLICANT: Hillan, Kenneth
;; APPLICANT: Kijavlin, Ivar
;; APPLICANT: Napier, Mary
;; APPLICANT: Roy, Margaret

```

; APPLICANT: Tumas, Daniel
; APPLICANT: Wood, William
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P2548P1C1
; CURRENT APPLICATION NUMBER: US/09/944,929
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 120
; SEQ ID NO 55
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-944-929-55

Query Match
Best Local Similarity 66.7%; Pred. No. 75;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QPGPVLSP 10
DB 658 QPGPLKKP 666

RESULT 10
US-10-174-590-60
; Sequence 60, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C42
; CURRENT APPLICATION NUMBER: US/10/174,590
; PRIOR FILING DATE: 2002-06-18
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 60
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-590-60

Query Match
Best Local Similarity 71.7%; Score 38; DB 9; Length 694;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QPGPVLSP 10
DB 658 QPGPLKKP 666

RESULT 11
US-10-176-758-60
; Sequence 60, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Godowski, Paul J.

```

```

; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C104
; CURRENT APPLICATION NUMBER: US/10/176,758
; PRIOR FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 60
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-758-60

Query Match
Best Local Similarity 71.7%; Score 38; DB 9; Length 694;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QPGPVLSP 10
DB 658 QPGPLKKP 666

RESULT 12
US-10-175-737-60
; Sequence 60, Application US/10175737
; Publication No. US20030013153A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C50
; CURRENT APPLICATION NUMBER: US/10/175,737
; PRIOR FILING DATE: 2002-06-19
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 60
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-737-60

Query Match
Best Local Similarity 71.7%; Score 38; DB 9; Length 694;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QPGPVLSP 10
DB 658 QPGPLKKP 666

RESULT 13
US-10-173-706-60
; Sequence 60, Application US/10173706
; Publication No. US2003002293A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian

```

```
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C7
; CURRENT APPLICATION NUMBER: US/10/173,706
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 60
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-706-60
```

```
Query Match          71.7%; Score 38; DB 9; Length 694;
Best Local Similarity 66.7%; Pred. No. 75;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 2 QPGPVLSP 10
Db 658 QPGPLKKP 666
```

```
RESULT 14
US-10-175-738-60
; Sequence 60, Application US/10175738
; Publication NO. US20030022294A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C45
; CURRENT APPLICATION NUMBER: US/10/175,738
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 60
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-738-60
```

```
Query Match          71.7%; Score 38; DB 9; Length 694;
Best Local Similarity 66.7%; Pred. No. 75;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 2 QPGPVLSP 10
Db 658 QPGPLKKP 666
```

```
RESULT 15
US-10-175-752-60
; Sequence 60, Application US/10175752
; Publication No. US20030022295A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C60
; CURRENT APPLICATION NUMBER: US/10/175,752
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 60
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-752-60
```

```
Query Match          71.7%; Score 38; DB 9; Length 694;
Best Local Similarity 66.7%; Pred. No. 75;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 2 QPGPVLSP 10
Db 658 QPGPLKKP 666
```

```
Search completed: February 16, 2003, 22:02:34
Job time : 2.197 secs
```

1

2

3

4

5

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 22:01:05 ; Search time 3.52891 Seconds
(without alignments)
1827.005 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_452

Perfect score: 53

Sequence: 1 ACPGVLSQP 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Pending_Patents_AA_Main:*

1: /cgn2_6/ptodata/2/paa/PCPUS_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US097_COMB.pep:*
22: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US099_COMB.pep:*
24: /cgn2_6/ptodata/2/paa/US100_COMB.pep:*
25: /cgn2_6/ptodata/2/paa/US101_COMB.pep:*
26: /cgn2_6/ptodata/2/paa/US102_COMB.pep:*
27: /cgn2_6/ptodata/2/paa/US60_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	100.0	685	27	US-60-243-468-1271
2	53	100.0	801	1	PCT-US01-08631-39827
3	53	100.0	843	1	PCT-US02-29964-410
4	53	100.0	852	21	US-09-763-909-2
5	53	100.0	865	1	PCT-US02-25829-214
6	43	81.1	71	1	PCT-US01-02723-214

7	43	81.1	71	21	US-09-707-351-214	Sequence 214, App
8	43	81.1	71	27	US-60-365-384-578	Sequence 578, App
9	42	79.2	348	1	PCT-US01-08631-44366	Sequence 254, App
10	42	79.2	348	1	PCT-US02-30474-2979	Sequence 44366, App
11	42	79.2	348	1	PCT-US02-30474-2979	Sequence 2979, App
12	42	79.2	348	27	US-60-324-631-2885	Sequence 2885, App
13	41	77.4	235	26	US-10-219-999-44244	Sequence 44244, A
14	41	77.4	250	21	US-09-708-427-54708	Sequence 54708, A
15	41	77.4	255	21	US-10-219-999-49842	Sequence 49842, A
16	41	77.4	310	26	US-10-219-999-38659	Sequence 38659, A
17	41	77.4	310	27	US-60-324-109-25624	Sequence 25624, A
18	41	77.4	316	21	US-09-708-427-54707	Sequence 54707, A
19	41	77.4	351	18	US-09-478-081-504	Sequence 504, App
20	41	77.4	353	19	US-09-513-996A-35642	Sequence 35642, A
21	41	77.4	353	21	US-09-708-427-1717	Sequence 1717, App
22	41	77.4	354	21	US-09-791-537-114482	Sequence 114482, A
23	41	77.4	357	26	US-10-219-999-51795	Sequence 51795, A
24	41	77.4	362	20	US-09-555-073-14	Sequence 14, App1
25	41	77.4	362	21	US-09-708-427-54706	Sequence 54706, A
26	41	77.4	362	26	US-60-324-109-18893	Sequence 3627, A
27	41	77.4	362	27	US-09-513-999-33627	Sequence 33627, A
28	41	77.4	365	19	US-09-513-996A-35641	Sequence 35641, A
29	41	77.4	365	19	US-09-573-655A-1205	Sequence 1205, App
30	41	77.4	365	19	US-09-573-655A-1205	Sequence 1205, App
31	41	77.4	365	21	US-09-708-427-1716	Sequence 1716, App
32	41	77.4	365	21	US-09-791-537-61865	Sequence 61865, App1
33	41	77.4	370	20	US-09-602-472-60	Sequence 60, App1
34	41	77.4	370	21	US-09-791-537-127939	Sequence 127939, A
35	41	77.4	397	19	US-09-513-996A-35640	Sequence 35640, A
36	41	77.4	397	21	US-09-708-427-1715	Sequence 1715, App
37	41	77.4	402	1	PCT-US02-27143-71	Sequence 71, App1
38	41	77.4	439	26	US-10-219-999-44281	Sequence 44281, A
39	41	77.4	439	27	US-60-324-109-24623	Sequence 24623, A
40	41	77.4	592	19	US-09-543-091-2	Sequence 2, App1
41	40	75.5	113	1	PCT-US01-01338-1677	Sequence 1677, App
42	40	75.5	113	21	US-09-764-877-1677	Sequence 1677, App
43	40	75.5	113	26	US-10-242-515-1677	Sequence 1677, App
44	40	75.5	199	1	PCT-US01-08631-50835	Sequence 50835, A
45	40	75.5	247	18	US-09-417-507-28769	Sequence 28769, A

ALIGNMENTS

RESULT 1
US-60-243-468-1271
Sequence 1271, Application US/60243468
GENERAL INFORMATION:
APPLICANT: Beasley, Ellen
TITLE OF INVENTION: ISOLATED HUMAN DRUG TARGET PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN DRUG TARGET PROTEINS,
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: C1000929
CURRENT APPLICATION NUMBER: US/60/243,468
CURRENT FILING DATE: 2000-10-27
NUMBER OF SEQ ID NOS: 2121
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1271
LENGTH: 685
TYPE: PRT
ORGANISM: HUMAN
US-60-243-468-1271

Query Match 100.0%; Score 53; DB 27; Length 685;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACPGVLSQP 10
DB 473 ACPGVLSQP 482
RESULT 2

PCT-US01-08631-39827
; Sequence 39827, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO: 39827
; LENGTH: 801
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-08631-39827

Query Match 100.0%; Score 53; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AOPGPVLSQP 10
DB 443 AOPGPVLSQP 452

RESULT 3
PCT-US02-29964-410
; Sequence 410, Application PC/TUS0229964
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhang, Jie
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Xue, Aidong J.
; APPLICANT: Wang, Dunrui
; APPLICANT: Ghosh, Malabika
; APPLICANT: Asundi, Vinod
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle W.
; APPLICANT: Weng, Gezhi
; APPLICANT: Haley-Vicente, Dana
; APPLICANT: Drmanac, Radoje T
; TITLE OF INVENTION: Novel Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 809ACIP PCT
; CURRENT APPLICATION NUMBER: PCT/US02/29964
; CURRENT FILING DATE: 2002-09-19
; PRIOR APPLICATION NUMBER: US 60/323,739
; PRIOR FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28

Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 992
; SOFTWARE: PC_FL_genes Version 6.0
; SEQ ID NO: 410
; LENGTH: 843
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-29964-410

Query Match 100.0%; Score 53; DB 1; Length 843;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AOPGPVLSQP 10
DB 485 AOPGPVLSQP 494

RESULT 4
US-09-763-909-2
; Sequence 2, Application US/09763909
; GENERAL INFORMATION:
; APPLICANT: Dikstein, Rivka
; APPLICANT: Yamle-Hezi, Avaya
; TITLE OF INVENTION: A TRANSCRIPTION FACTOR TFIIID SUBUNIT,
; TITLE OF INVENTION: TAFII105, POLYPEPTIDES, DNA ENCODING THEREFOR AND
; FILE REFERENCE: 13005/002001
; CURRENT APPLICATION NUMBER: US/09/763,909
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 2
; LENGTH: 852
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-763-909-2

Query Match 100.0%; Score 53; DB 21; Length 852;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AOPGPVLSQP 10
DB 443 AOPGPVLSQP 452

RESULT 5
PCT-US02-25829-25
; Sequence 25, Application PC/TUS0225829
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: BAROSSO, Ines
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: BECHA, Shanya D.
; APPLICANT: BLAKE, Julie J.
; APPLICANT: BOROWSKY, Mark L.
; APPLICANT: BUREFORD, Neil
; APPLICANT: DUGGAN, Brendan M.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: EMERLING, Brooke M.
; APPLICANT: FORSTHE, Ian J.
; APPLICANT: GIETZEN, Kimberly J.
; APPLICANT: GORVAD, Ann E.
; APPLICANT: GRIFFIN, Jennifer A.
; APPLICANT: HAPALIA, April J.A.
; APPLICANT: HONCHELL, Cynthia D.
; APPLICANT: ISON, Craig H.
; APPLICANT: KHAN, Farrah A.
; APPLICANT: LAL, Preeti G.
; APPLICANT: LEE, Ernestine A.
; APPLICANT: LEE, Sally
; APPLICANT: LEE, Soo Yeun

```

: APPLICANT: LI, Joana X.
: APPLICANT: LU, Yung Aina M.
: APPLICANT: LU, Yan
: APPLICANT: LEHR-MASON, Patricia M.
: APPLICANT: NGUYEN, Daniel B.
: APPLICANT: RAMKUMAR, Jayalaxmi
: APPLICANT: SPRAGUE, William W.
: APPLICANT: TANG, Y. Tom
: APPLICANT: THANGAVELU, Kavitha
: APPLICANT: THORNTON, Michael
: APPLICANT: TRAN, Uyen K.
: APPLICANT: WALIA, Narinder K.
: APPLICANT: WARREN, Bridget A.
: APPLICANT: XU, Yuming
: APPLICANT: YAO, Monique G.
: APPLICANT: YUE, Henry
: APPLICANT: YUE, Huijin
: APPLICANT: ZEBARADIAN, Yeganeh
: TITLE OF INVENTION: NUCLEIC ACID ASSOCIATED PROTEINS
: FILE REFERENCE: PF-1146 PCT
: CURRENT APPLICATION NUMBER: PCT/US02/25829
: CURRENT FILING DATE: 2002-08-14
: PRIOR APPLICATION NUMBER: US 60/313,111
: PRIOR FILING DATE: 2001-08-17
: PRIOR APPLICATION NUMBER: US 60/314,682
: PRIOR FILING DATE: 2001-08-24
: PRIOR APPLICATION NUMBER: US 60/314,756
: PRIOR FILING DATE: 2001-08-24
: PRIOR APPLICATION NUMBER: US 60/315,105
: PRIOR FILING DATE: 2001-08-27
: PRIOR APPLICATION NUMBER: US 60/316,751
: PRIOR FILING DATE: 2001-08-31
: PRIOR APPLICATION NUMBER: US 60/316,856
: PRIOR FILING DATE: 2001-08-31
: PRIOR APPLICATION NUMBER: US 60/328,185
: PRIOR FILING DATE: 2001-10-05
: NUMBER OF SEQ ID NOS: 66
: SOFTWARE: PERL Program
: SEQ ID NO 25
: LENGTH: 865
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: OTHER INFORMATION: Incyte ID No: 4398735CD1
PCT-US02-25829-25
Query Match 100.0%; Score 53; DB 1; Length 865;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AOPGPVLSOP 10
Db 512 AOPGPVLSOP 521
RESULT 6
PCT-US01-02723-214
: Sequence 214, Application PC/TUS0102723
: GENERAL INFORMATION:
: APPLICANT: Hyseq, Inc.
: APPLICANT: Yeung, George
: APPLICANT: Ford, John E.
: APPLICANT: Boyle, Bryan J.
: APPLICANT: Arterburn, Matthew C.
: APPLICANT: Dirmanc, Radoje T.
: APPLICANT: Tang, Y. Tom
: APPLICANT: Liu, Chenghua
: APPLICANT: Asundi, Vinod
: APPLICANT: Zhou, Ping
: APPLICANT: Wehrman, Thomas
: TITLE OF INVENTION: NOVEL FETAL NUCLEIC ACIDS AND POLYPEPTIDES
: FILE REFERENCE: 21272-020
```

```

: CURRENT APPLICATION NUMBER: PCT/US01/02723
: CURRENT FILING DATE: 2001-01-25
: PRIOR APPLICATION NUMBER: 09/707,351
: PRIOR FILING DATE: 2000-11-06
: PRIOR APPLICATION NUMBER: 09/491,404
: PRIOR FILING DATE: 2000-01-25
: PRIOR APPLICATION NUMBER: 09/663,870
: PRIOR FILING DATE: 2000-09-15
: NUMBER OF SEQ ID NOS: 1262
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 214
: LENGTH: 71
: TYPE: PRT
: ORGANISM: homo sapiens
PCT-US01-02723-214
Query Match 81.1%; Score 43; DB 1; Length 71;
Best Local Similarity 70.0%; Pred. No. 41;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
OY 1 AOPGPVLSOP 10
Db 40 SRPGVLSOP 49
RESULT 7
US-09-707-351-214
: Sequence 214, Application US/09707351
: GENERAL INFORMATION:
: APPLICANT: Yeung, George
: APPLICANT: Ford, John E.
: APPLICANT: Boyle, Bryan J.
: APPLICANT: Arterburn, Matthew C.
: APPLICANT: Dirmanc, Radoje T.
: APPLICANT: Tang, Y. Tom
: APPLICANT: Liu, Chenghua
: APPLICANT: Asundi, Vinod
: APPLICANT: Zhou, Ping
: APPLICANT: Wehrman, Thomas
: TITLE OF INVENTION: NOVEL FETAL NUCLEIC ACIDS AND POLYPEPTIDES
: FILE REFERENCE: 796
: CURRENT APPLICATION NUMBER: US/09/707,351
: CURRENT FILING DATE: 2000-11-06
: PRIOR APPLICATION NUMBER: 09/663,870
: PRIOR FILING DATE: 2000-09-15
: PRIOR APPLICATION NUMBER: 09/491,404
: PRIOR FILING DATE: 2000-01-25
: NUMBER OF SEQ ID NOS: 954
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 214
: LENGTH: 71
: TYPE: PRT
: ORGANISM: homo sapiens
US-09-707-351-214
Query Match 81.1%; Score 43; DB 21; Length 71;
Best Local Similarity 70.0%; Pred. No. 41;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
OY 1 AOPGPVLSOP 10
Db 40 SRPGVLSOP 49
RESULT 8
US-60-365-384-578
: Sequence 578, Application US/60365384
: GENERAL INFORMATION:
: APPLICANT: Tang, Y. Tom
: APPLICANT: Wang, Jian-Tui
: APPLICANT: Wang, Zhiwei
: APPLICANT: Zhang, Jie
: APPLICANT: Zhou, Ping
```

```
APPLICANT: Wehrman, Tom
APPLICANT: Wang, Jian-Rui
APPLICANT: Ghosh, Malabika
APPLICANT: Zhao, Qing A.
APPLICANT: Asundi, Vinod
APPLICANT: Ren, Feiyan
APPLICANT: Xue, Aidong J.
APPLICANT: Ma, Yungqing
APPLICANT: Wang, Dunrui
APPLICANT: Weng, Gezhi
TITLE OF INVENTION: Novel Nucleic Acids and
FILE REFERENCE: 814
CURRENT APPLICATION NUMBER: US/60/365,384
CURRENT FILING DATE: 2002-03-14
NUMBER OF SEQ ID NOS: 666
SOFTWARE: PC_FL_genes Version 6.0
SEQ ID NO 578
LENGTH: 71
TYPE: PRT
ORGANISM: Homo sapiens
US-60-365-384-578
```

```
Query Match
Best Local Similarity 81.1%; Score 43; DB 27; Length 71;
Best Local Similarity 70.0%; Pred. No. 41;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 AOPGPVLSOP 10
Db 40 SRGPVLSER 49

RESULT 9
US-60-365-384-254
; Sequence 254, Application US/60365384
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Zhang, Jie
; APPLICANT: Zhou, Ping
; APPLICANT: Wehrman, Tom
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Ghosh, Malabika
; APPLICANT: Zhao, Qing A.
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Xue, Aidong J.
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Weng, Gezhi
TITLE OF INVENTION: Novel Nucleic Acids and
FILE REFERENCE: 814
CURRENT APPLICATION NUMBER: US/60/365,384
CURRENT FILING DATE: 2002-03-14
NUMBER OF SEQ ID NOS: 666
SOFTWARE: PC_FL_genes Version 6.0
SEQ ID NO 254
LENGTH: 91
TYPE: PRT
ORGANISM: Homo sapiens
US-60-365-384-254
```

```
Query Match
Best Local Similarity 81.1%; Score 43; DB 27; Length 91;
Best Local Similarity 70.0%; Pred. No. 54;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
```

```
RESULT 10
PCT-US01-08631-44366
; Sequence 44366, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hysq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 44366
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(348)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
PCT-US01-08631-44366
```

```
Query Match
Best Local Similarity 79.2%; Score 42; DB 1; Length 348;
Best Local Similarity 80.0%; Pred. No. 3; Seq+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 AOPGPVLSOP 10
Db 99 AOPGPVLSER 108
```

```
RESULT 11
PCT-US02-30474-2979
; Sequence 2979, Application PC/TUS0230474
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Xue, Aidong J.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Ghosh, Malabika
; APPLICANT: Wang, Dunrui
; APPLICANT: Ma, Yungqing
; APPLICANT: Asundi, Vinod
; APPLICANT: Wang, Zhiwei
; APPLICANT: Weng, Gezhi
; APPLICANT: Haley-Vicente, Dana
; APPLICANT: Drmanac, Radolje T
TITLE OF INVENTION: Novel Nucleic Acids and
FILE REFERENCE: 810CIP PCT
CURRENT APPLICATION NUMBER: PCT/US02/30474
CURRENT FILING DATE: 2002-09-24
PRIOR APPLICATION NUMBER: US 60/324,631
PRIOR FILING DATE: 2001-09-24
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02623
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03
```


;; PRIOR APPLICATION NUMBER: US 09/560,875
;; PRIOR FILING DATE: 2000-04-27
;; PRIOR APPLICATION NUMBER: PCT/US01/03800
;; PRIOR FILING DATE: 2001-02-05
;; PRIOR APPLICATION NUMBER: US 09/515,126
;; PRIOR FILING DATE: 2000-02-28
;; Remaining prior application data removed - See file Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 3476
;; SOFTWARE: PL_FL_genes Version 6.0
;; SEQ ID NO 2979
;; LENGTH: 348
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)...(348)
;; OTHER INFORMATION: Xaa - any amino acid or symbol as shown in the table 8 as set for
PCT-US02-30474-2379

Query Match 79.2%; Score 42; DB 1; Length 348;
Best Local Similarity 80.0%; Pred. No. 3.3e+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AOPGPVLSOP 10
Db 99 AOPGPVLSVP 108

RESULT 12
US-60-324-631-2885

;; Sequence 2885, Application US/60324631
;; GENERAL INFORMATION:

;; APPLICANT: Tang, Y. Tom
;; APPLICANT: Zhang, Jie

;; APPLICANT: Ren, Feiyan
;; APPLICANT: Xue, Aiqiong J.

;; APPLICANT: Zhao, Qing A.
;; APPLICANT: Wang, Jian-Rui

;; APPLICANT: Mehrman, Tom
;; APPLICANT: Zhou, Ping

;; APPLICANT: Ghosh, Malabika
;; APPLICANT: Wang, Dunrui

;; APPLICANT: Ma, Yungqing
;; APPLICANT: Asundi, Vinod

;; APPLICANT: Wang, Zhiwei
;; APPLICANT: Meng, Gezhi

;; APPLICANT: Haley-Vicente, Dana
;; APPLICANT: Dimaucac, Radoje T

;; TITLE OF INVENTION: Novel Nucleic Acids and
;; FILE REFERENCE: 810

;; CURRENT APPLICATION NUMBER: US/60/324,631
;; CURRENT FILING DATE: 2001-09-24

;; PRIOR APPLICATION NUMBER: US 09/488,725
;; PRIOR FILING DATE: 2000-01-21

;; PRIOR APPLICATION NUMBER: US 09/552,317
;; PRIOR FILING DATE: 2000-04-25

;; PRIOR APPLICATION NUMBER: PCT/US00/35017
;; PRIOR FILING DATE: 2000-12-22

;; PRIOR APPLICATION NUMBER: US 09/491,404
;; PRIOR FILING DATE: 2000-01-25

;; PRIOR APPLICATION NUMBER: PCT/US01/02623
;; PRIOR FILING DATE: 2001-01-25

;; PRIOR APPLICATION NUMBER: US 09/496,914
;; PRIOR FILING DATE: 2000-02-03

;; PRIOR APPLICATION NUMBER: US 09/560,875
;; PRIOR FILING DATE: 2000-04-27

;; PRIOR APPLICATION NUMBER: PCT/US01/03800
;; PRIOR FILING DATE: 2001-02-05

;; PRIOR APPLICATION NUMBER: US 09/515,126
;; PRIOR FILING DATE: 2000-02-28

;; PRIOR APPLICATION NUMBER: US 09/577,409
;; PRIOR FILING DATE: 2000-05-18

;; PRIOR APPLICATION NUMBER: PCT/US01/04927
;; PRIOR FILING DATE: 2001-02-26
;; PRIOR APPLICATION NUMBER: US 09/519,705
;; PRIOR FILING DATE: 2000-03-07
;; PRIOR APPLICATION NUMBER: US 09/574,454
;; PRIOR FILING DATE: 2000-05-19
;; PRIOR APPLICATION NUMBER: PCT/US01/04941
;; PRIOR FILING DATE: 2001-03-05
;; PRIOR APPLICATION NUMBER: US 09/540,217
;; PRIOR FILING DATE: 2000-03-31
;; PRIOR APPLICATION NUMBER: US 09/649,167
;; PRIOR FILING DATE: 2000-08-23
;; PRIOR APPLICATION NUMBER: PCT/US01/08631
;; PRIOR FILING DATE: 2001-03-30
;; PRIOR APPLICATION NUMBER: US 09/552,929
;; PRIOR FILING DATE: 2000-04-18
;; PRIOR APPLICATION NUMBER: US 09/770,160
;; PRIOR FILING DATE: 2001-01-26
;; PRIOR APPLICATION NUMBER: PCT/US01/08656
;; PRIOR FILING DATE: 2001-04-18
;; PRIOR APPLICATION NUMBER: US 09/577,408
;; PRIOR FILING DATE: 2000-05-18
;; PRIOR APPLICATION NUMBER: PCT/US01/14827
;; PRIOR FILING DATE: 2001-05-16
;; NUMBER OF SEQ ID NOS: 3334
;; SOFTWARE: PL_FL_genes Version 6.0
;; SEQ ID NO 2885
;; LENGTH: 348
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)...(348)
;; OTHER INFORMATION: Xaa - any amino acid or symbol as shown in the table 8 as set
US-60-324-631-2885

Query Match 79.2%; Score 42; DB 27; Length 348;
Best Local Similarity 80.0%; Pred. No. 3.3e+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AOPGPVLSOP 10
Db 99 AOPGPVLSVP 108

RESULT 13
US-10-219-999-44244

;; Sequence 44244, Application US/10219999
;; GENERAL INFORMATION:

;; APPLICANT: Cao, Yongwei
;; APPLICANT: Edgerton, Michael D

;; APPLICANT: Hinkle, Gregory J.
;; APPLICANT: Kovalic, David K.

;; APPLICANT: Liu, Jingdong
;; APPLICANT: Stein, Joshua

;; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT
;; FILE REFERENCE: 38-10452726YC

;; CURRENT APPLICATION NUMBER: US/10/219,999
;; CURRENT FILING DATE: 2002-08-15

;; PRIOR APPLICATION NUMBER: US 60/324,109
;; PRIOR FILING DATE: 2001-09-21

;; PRIOR APPLICATION NUMBER: US 60/312,544
;; PRIOR FILING DATE: 2001-08-15

;; NUMBER OF SEQ ID NOS: 63520
;; SEQ ID NO 44244

;; LENGTH: 235
;; TYPE: PRT
;; ORGANISM: Zea mays

US-10-219-999-44244

Query Match 77.4%; Score 41; DB 26; Length 235;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AOPGPVLS 8
 Db 23 AOPGPVLS 30

Search completed: February 16, 2003, 22:18:32
 Job time : 4.52891 secs

RESULT 14

US-09-708-427-54708
 ; Sequence 54708, Application US/09708427
 ; GENERAL INFORMATION:
 ; APPLICANT: N. ALEXANDROV et al.
 ; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES.
 ; FILE REFERENCE: 2750-1243P
 ; CURRENT APPLICATION NUMBER: US/09/708,427
 ; CURRENT FILING DATE: 2000-11-09
 ; NUMBER OF SEQ ID NOS: 85364
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 54708
 ; LENGTH: 250
 ; TYPE: PRT
 ; ORGANISM: Zea mays subsp. mays
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: 1..250
 ; OTHER INFORMATION: Xaa is any amino acid
 ; NAME/KEY: misc_feature
 ; LOCATION: 1..250
 ; OTHER INFORMATION: Ceres Seq. ID 1933788
 US-09-708-427-54708

Query Match

Best Local Similarity 77.4%; Score 41; DB 21; Length 250;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AOPGPVLS 8
 Db 48 AOPGPVLS 55

RESULT 15

US-10-219-999-49842
 ; Sequence 49842, Application US/10219999
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Edgerton, Michael D
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Kovacic, David K.
 ; APPLICANT: Liu, Jiongong
 ; APPLICANT: Stein, Joshua
 ; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT
 ; FILE REFERENCE: 38-10(52726)C
 ; CURRENT APPLICATION NUMBER: US/10/219,999
 ; CURRENT FILING DATE: 2002-08-15
 ; PRIOR APPLICATION NUMBER: US 60/324,109
 ; PRIOR FILING DATE: 2001-09-21
 ; PRIOR APPLICATION NUMBER: US 60/312,544
 ; PRIOR FILING DATE: 2001-08-15
 ; NUMBER OF SEQ ID NOS: 63520
 ; SEQ ID NO 49842
 ; LENGTH: 255
 ; TYPE: PRT
 ; ORGANISM: Zea mays
 US-10-219-999-49842

Query Match

Best Local Similarity 77.4%; Score 41; DB 26; Length 255;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AOPGPVLS 8
 Db 56 AOPGPVLS 63

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 22:00:19 ; Search time 0.406852 Seconds
(without alignments)
2076.603 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_452

Perfect score: 53

Sequence: 1 AQPGLVLSQP 10

Scoring table:

BL0SUM62

Gapop 10.0 , Gapext 0.5

Searched: 429898 seqs, 84487048 residues

Total number of hits satisfying chosen parameters: 429898

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	44	83.0	136	1 PCT-US02-32727-6510	Sequence 6510, Ap
2	44	83.0	136	US-10-057-498-6510	Sequence 6510, Ap
3	41	77.4	370	US-09-602-472A-60	Sequence 60, Appl
4	39	73.6	775	1 PCT-US02-10366-58	Sequence 58, Appl
5	38	71.7	608	6 US-10-351-334-213	Sequence 213, App
6	38	71.7	694	6 US-10-125-923A-60	Sequence 60, Appl
7	38	71.7	694	6 US-10-205-892-60	Sequence 60, Appl
8	38	71.7	694	6 US-10-174-575-60	Sequence 60, Appl
9	38	71.7	694	6 US-10-174-575A-60	Sequence 60, Appl
10	38	71.7	694	6 US-10-187-755-60	Sequence 60, Appl
11	38	71.7	694	6 US-10-187-749-60	Sequence 60, Appl
12	38	71.7	694	6 US-10-199-672-60	Sequence 60, Appl
13	38	71.7	694	6 US-10-194-486-60	Sequence 60, Appl
14	38	71.7	716	6 US-10-351-334-370	Sequence 370, App
15	37	69.8	194	6 US-10-299-003-52	Sequence 52, Appl
16	37	69.8	222	5 US-09-724-676-69050	Sequence 69050, A
17	37	69.8	222	5 US-09-724-676-69051	Sequence 69051, A
18	37	69.8	222	5 US-09-724-676A-69050	Sequence 69050, A
19	37	69.8	222	5 US-09-724-676A-69051	Sequence 69051, A
20	37	69.8	234	5 US-09-724-676-69052	Sequence 69052, A
21	37	69.8	234	5 US-09-724-676A-69052	Sequence 69052, A
22	37	69.8	333	5 US-09-724-676-64147	Sequence 64147, A
23	37	69.8	333	5 US-09-724-676A-64147	Sequence 64147, A
24	37	69.8	559	6 US-10-289-003-15	Sequence 15, Appl
25	37	69.8	756	6 US-10-338-462-9	Sequence 9, Appl
26	37	69.8	768	5 US-09-724-676-69047	Sequence 69047, A

27	37	69.8	768	5 US-09-724-676A-69047	Sequence 69047, A
28	37	69.8	967	6 US-10-017-161-1766	Sequence 1766, Ap
29	37	69.8	972	6 US-10-276-781-1513	Sequence 1513, Ap
30	37	69.8	972	6 US-10-276-781-1535	Sequence 1535, Ap
31	36	67.9	183	6 US-10-264-237-2472	Sequence 2472, Ap
32	36	67.9	190	6 US-10-203-138A-13166	Sequence 13166, A
33	36	67.9	362	5 US-09-373-902-7	Sequence 7, Appl
34	36	67.9	422	6 US-10-264-237-2090	Sequence 2090, Ap
35	36	67.9	1109	1 PCT-US02-40059-8	Sequence 8, Appl
36	36	67.9	1220	6 US-10-144-779-525	Sequence 525, App
37	36	67.9	2507	1 PCT-US02-40059-2	Sequence 2, Appl
38	35	66.0	53	6 US-10-203-138A-11392	Sequence 11392, A
39	35	66.0	72	1 PCT-US02-32727-23821	Sequence 23821, A
40	35	66.0	72	6 US-10-057-498-23821	Sequence 23821, A
41	35	66.0	172	1 PCT-US02-32850A-5	Sequence 5, Appl
42	35	66.0	172	1 PCT-US02-32850A-5	Sequence 5, Appl
43	35	66.0	350	1 PCT-US02-32874-3	Sequence 3, Appl
44	35	66.0	350	5 US-09-906-777B-236	Sequence 236, App
45	35	66.0	350	5 US-09-904-011C-236	Sequence 236, App

ALIGNMENTS

RESULT 1
PCT-US02-32727-6510
Sequence 6510, Application PC/TUS0232727

GENERAL INFORMATION:

APPLICANT: Mitcham, Jennifer
APPLICANT: Skeiky, Yasir
APPLICANT: Persing, David
APPLICANT: Bhatia, Ajay
APPLICANT: Malsomeuve, Jean Francois
APPLICANT: Zhang, Yanni
APPLICANT: Wang, Siqing
APPLICANT: Jen, Shiyian
APPLICANT: Lodes, Michael
APPLICANT: Benson, Darin
APPLICANT: Jones, Robert
APPLICANT: Carter, Darick
APPLICANT: Barth, Brenda
APPLICANT: Douglass, John
TITLE OF INVENTION: Compositions and Methods for the therapy and Diagnosis of Acne
FILE REFERENCE: 210121.514C1
CURRENT APPLICATION NUMBER: PCT/US02/32727
CURRENT FILING DATE: 2002-10-11
NUMBER OF SEQ ID NOS: 30992
SEQ ID NO 6510
LENGTH: 136
TYPE: PRT
ORGANISM: Propionibacterium

PCT-US02-32727-6510

Query Match

Best local Similarity 83.0%; Score 44; DB 1; Length 136;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 61 AQPGLVLSQP 10

US-10-057-498-6510

Sequence 6510, Application US/10057498

GENERAL INFORMATION:

APPLICANT: Mitcham, Jennifer

APPLICANT: Skeiky, Yasir

APPLICANT: Persing, David

TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acne

FILE REFERENCE: 210121.514

CURRENT APPLICATION NUMBER: US/10/057,498

CURRENT FILING DATE: 2001-04-20

NUMBER OF SEQ ID NOS: 29212
; SEQ ID NO 6510
; LENGTH: 136
; TYPE: PRT
; ORGANISM: Propionibacterium
US-10-057-498-6510

Query Match
Best Local Similarity 83.0%; Score 44; DB 6; Length 136;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 1 ACPGVLSQ 10
Db 61 ACPGVLSQ 70

RESULT 3
US-09-602-472A-60
; Sequence 60, Application US/09602472A
; GENERAL INFORMATION:
; APPLICANT: Bidney, Dennis L.
; APPLICANT: Hu, Xu
; APPLICANT: Li, Guohua
; TITLE OF INVENTION: Sunflower Disease Resistance Genes
; FILE REFERENCE: 35718/200630
; CURRENT APPLICATION NUMBER: US/09/602,472A
; CURRENT FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/140,876
; PRIOR FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 60
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Lycopersicon esculentum
US-09-602-472A-60

Query Match
Best Local Similarity 77.4%; Score 41; DB 5; Length 370;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ACPGVLSQ 8
Db 161 ACPGVLSQ 168

RESULT 4
PCT-US02-10366-58
; Sequence 58, Application PC/TUS0210366
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Murallidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter
; APPLICANT: Smithson, Glennda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zernusen, Bryan
; APPLICANT: Tchernev, Vellizar
; APPLICANT: Gangoli, Eshe
; APPLICANT: Vernet, Corine
; APPLICANT: Spytek, Kimberly
; APPLICANT: Malvankar, Uriel
; APPLICANT: Patursajan, Meera
; APPLICANT: Miller, Charles
; APPLICANT: Taupier, Raymond J. Jr.
; APPLICANT: Heyes, Melvyn
; APPLICANT: Ju, Jiongfang
; APPLICANT: Peyman, John

APPLICANT: Catterton, Elina
APPLICANT: MacDougall, John
APPLICANT: Edinger, Shlomit
APPLICANT: Stone, David
APPLICANT: Mazur, Ann
; TITLE OF INVENTION: NOVEL ANTIBODIES THAT BIND TO ANTIGENIC POLYPEPTIDES,
; TITLE OF INVENTION: NUCLEIC
; FILE REFERENCE: 21402-322A-061
; CURRENT APPLICATION NUMBER: PCT/US02/10366
; CURRENT FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 10/114153
; PRIOR FILING DATE: 2002-04-02
; PRIOR APPLICATION NUMBER: 60/281086
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/281906
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/282020
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/282930
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 60/283512
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/283444
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/283657
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/283710
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/283678
; PRIOR FILING DATE: 2001-04-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 251
; SEQ ID NO 58
; LENGTH: 775
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-10366-58

Query Match
Best Local Similarity 73.6%; Score 39; DB 1; Length 775;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 ACPGVLSQ 9
Db 264 ACPGVLSQ 272

RESULT 5
US-10-351-334-213
; Sequence 213, Application US/10351334
; GENERAL INFORMATION:
; APPLICANT: Komatsoulis et al
; TITLE OF INVENTION: 98 Human Secreted Proteins
; FILE REFERENCE: P2031P2
; CURRENT APPLICATION NUMBER: US/10/351,334
; CURRENT FILING DATE: 2003-01-27
; PRIOR APPLICATION NUMBER: 60/350,898
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: 09/489,847
; PRIOR FILING DATE: 2000-01-24
; PRIOR APPLICATION NUMBER: PCT/US99/17130
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: 60/094,657
; PRIOR FILING DATE: 1998-07-30
; PRIOR APPLICATION NUMBER: 60/095,486
; PRIOR FILING DATE: 1998-08-05
; PRIOR APPLICATION NUMBER: 60/096,319
; PRIOR FILING DATE: 1998-08-12
; PRIOR APPLICATION NUMBER: 60/095,454
; PRIOR FILING DATE: 1998-08-06
; PRIOR APPLICATION NUMBER: 60/095,455
; PRIOR FILING DATE: 1998-08-06

```

; NUMBER OF SEQ ID NOS: 376
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 213
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (265)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (397)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-351-334-213
```

```

Query Match          71.7%; Score 38; DB 6; Length 608;
Best Local Similarity 66.7%; Pred. No. 2.3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      2 QPGPVLSOP 10
      |||||:|:|
Db      353 QPGPLKKP 361
```

```

RESULT 6
US-10-125-923A-60
; Sequence 60, Application US/10125923A
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jjian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C79
; CURRENT APPLICATION NUMBER: US/10/125,923A
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 10/052586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 60
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-125-923A-60
```

```

Query Match          71.7%; Score 38; DB 6; Length 694;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      2 QPGPVLSOP 10
      |||||:|:|
Db      658 QPGPLKKP 666
```

```

RESULT 7
US-10-205-892-60
; Sequence 60, Application US/10205892
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jjian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C517
; CURRENT APPLICATION NUMBER: US/10/205,892
; PRIOR FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: 10/052586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 60
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-205-892-60

Query Match          71.7%; Score 38; DB 6; Length 694;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```

RESULT 8
US-10-174-575-60
; Sequence 60, Application US/10174575
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jjian
; APPLICANT: Desnoyers, Luc
```

APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C35
CURRENT APPLICATION NUMBER: US/10/174,575
CURRENT FILING DATE: 2002-06-18
PRIOR APPLICATION NUMBER: 10/052586
PRIOR FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
Remaining Prior Application data removed - See file wrapper or PALM.
SEQU ID NO 60
LENGTH: 694
TYPE: PRT
ORGANISM: Homo Sapien
US-10-174-575-60
Query Match
Best Local Similarity 71.7%; Score 38; DB 6; Length 694;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
OY 2 QPGVLSQP 10
DB 658 QPGPLKRP 666
RESULT 9
US-10-174-575A-60
Sequence 60, Application US/10174575A
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C35
CURRENT APPLICATION NUMBER: US/10/174,575A
CURRENT FILING DATE: 2002-06-18
PRIOR APPLICATION NUMBER: 10/052586
PRIOR FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263

PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
Remaining Prior Application data removed - See file wrapper or PALM.
SEQU ID NO 60
LENGTH: 694
TYPE: PRT
ORGANISM: Homo Sapien
US-10-174-575A-60
Query Match
Best Local Similarity 71.7%; Score 38; DB 6; Length 694;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
OY 2 QPGVLSQP 10
DB 658 QPGPLKRP 666
RESULT 10
US-10-187-755-60
Sequence 60, Application US/10187755
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C1
CURRENT APPLICATION NUMBER: US/10/187,755
CURRENT FILING DATE: 2002-07-01
PRIOR APPLICATION NUMBER: US/10/052,586
PRIOR FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544

;; PRIOR FILING DATE: 1997-10-28
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 612
;; SEQ ID NO 60
;; LENGTH: 694
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-187-755-60

Query Match 71.7%; Score 38; DB 6; Length 694;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 QPGPVLSSOP 10
|||||:|
DB 658 QPGPLKKP 666

RESULT 11
US-10-187-749-60
;; Sequence 60, Application US/10187749
;; GENERAL INFORMATION:
;; APPLICANT: Baker, Kevin P.
;; APPLICANT: Chen, Jian
;; APPLICANT: Desnoyers, Luc
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Pan, James
;; APPLICANT: Smith, Victoria
;; APPLICANT: Watanabe, Colin K.
;; APPLICANT: Zhang, Zemin
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P3430R1C1
;; CURRENT APPLICATION NUMBER: US/10/187,749
;; CURRENT FILING DATE: 2002-07-01
;; PRIOR APPLICATION NUMBER: US/10/052,586
;; PRIOR FILING DATE: 2002-01-15
;; PRIOR APPLICATION NUMBER: 60/059263
;; PRIOR FILING DATE: 1997-09-18
;; PRIOR APPLICATION NUMBER: 60/059266
;; PRIOR FILING DATE: 1997-09-18
;; PRIOR APPLICATION NUMBER: 60/062250
;; PRIOR FILING DATE: 1997-10-17
;; PRIOR APPLICATION NUMBER: 60/063120
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/063121
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/063486
;; PRIOR FILING DATE: 1997-10-21
;; PRIOR APPLICATION NUMBER: 60/063540
;; PRIOR FILING DATE: 1997-10-28
;; PRIOR APPLICATION NUMBER: 60/063541
;; PRIOR FILING DATE: 1997-10-28
;; PRIOR APPLICATION NUMBER: 60/063544
;; PRIOR FILING DATE: 1997-10-28
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 612
;; SEQ ID NO 60
;; LENGTH: 694
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-187-749-60

Query Match 71.7%; Score 38; DB 6; Length 694;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 QPGPVLSSOP 10
|||||:|
DB 658 QPGPLKKP 666

RESULT 12
US-10-199-672-60
;; Sequence 60, Application US/10199672
;; GENERAL INFORMATION:
;; APPLICANT: Baker, Kevin P.
;; APPLICANT: Chen, Jian
;; APPLICANT: Desnoyers, Luc
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Pan, James
;; APPLICANT: Smith, Victoria
;; APPLICANT: Watanabe, Colin K.
;; APPLICANT: Wood, William I.
;; APPLICANT: Zhang, Zemin
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P3430R1C1
;; CURRENT APPLICATION NUMBER: US/10/199,672
;; CURRENT FILING DATE: 2002-07-18
;; PRIOR APPLICATION NUMBER: US/10/052,586
;; PRIOR FILING DATE: 2002-01-15
;; PRIOR APPLICATION NUMBER: 60/059263
;; PRIOR FILING DATE: 1997-09-18
;; PRIOR APPLICATION NUMBER: 60/059266
;; PRIOR FILING DATE: 1997-09-18
;; PRIOR APPLICATION NUMBER: 60/062250
;; PRIOR FILING DATE: 1997-10-17
;; PRIOR APPLICATION NUMBER: 60/063120
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/063121
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/063486
;; PRIOR FILING DATE: 1997-10-21
;; PRIOR APPLICATION NUMBER: 60/063540
;; PRIOR FILING DATE: 1997-10-28
;; PRIOR APPLICATION NUMBER: 60/063541
;; PRIOR FILING DATE: 1997-10-28
;; PRIOR APPLICATION NUMBER: 60/063544
;; PRIOR FILING DATE: 1997-10-28
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 612
;; SEQ ID NO 60
;; LENGTH: 694
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-199-672-60

Query Match 71.7%; Score 38; DB 6; Length 694;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 QPGPVLSSOP 10
|||||:|
DB 658 QPGPLKKP 666

RESULT 13
US-10-194-486-60
;; Sequence 60, Application US/10194486
;; GENERAL INFORMATION:
;; APPLICANT: Baker, Kevin P.
;; APPLICANT: Chen, Jian
;; APPLICANT: Desnoyers, Luc
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Pan, James
;; APPLICANT: Smith, Victoria
;; APPLICANT: Watanabe, Colin K.
;; APPLICANT: Wood, William I.

APPLICANT: Zhang Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C1
CURRENT APPLICATION NUMBER: US/10/194,486
CURRENT FILING DATE: 2002-07-11
PRIOR APPLICATION NUMBER: US/10/052,586
PRIOR FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
Remaining Prior Application data removed - See File Wrapper or PALM.
SEQ ID NO 60
LENGTH: 694
TYPE: PRT
ORGANISM: Homo Sapien
US-10-194-486-60

Query Match
Best Local Similarity 71.7%; Score 38; DB 6; Length 694;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QGPEVLSOP 10
DB 658 QGPEVLSKP 666

RESULT 14
US-10-351-334-370
Sequence 370, Application US/10351334
GENERAL INFORMATION:
APPLICANT: Komatsoulis et al
TITLE OF INVENTION: 98 Human Secreted Proteins
FILE REFERENCE: P2031P2
CURRENT APPLICATION NUMBER: US/10/351,334
CURRENT FILING DATE: 2003-01-27
PRIOR APPLICATION NUMBER: 60/350,898
PRIOR FILING DATE: 2002-01-25
PRIOR APPLICATION NUMBER: 09/489,847
PRIOR FILING DATE: 2000-01-24
PRIOR APPLICATION NUMBER: PCT/US99/17130
PRIOR FILING DATE: 1999-07-29
PRIOR APPLICATION NUMBER: 60/094,657
PRIOR FILING DATE: 1998-07-30
PRIOR APPLICATION NUMBER: 60/095,486
PRIOR FILING DATE: 1998-08-05
PRIOR APPLICATION NUMBER: 60/096,319
PRIOR FILING DATE: 1998-08-12
PRIOR APPLICATION NUMBER: 60/095,454
PRIOR FILING DATE: 1998-08-06
PRIOR APPLICATION NUMBER: 60/095,455
PRIOR FILING DATE: 1998-08-06
NUMBER OF SEQ ID NOS: 376
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 370
LENGTH: 716
TYPE: PRT

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (2)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (373)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (705)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-351-334-370

Query Match
Best Local Similarity 71.7%; Score 38; DB 6; Length 716;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QGPEVLSOP 10
DB 461 QGPEVLSKP 469

RESULT 15
US-10-299-003-52
Sequence 52, Application US/10299003
GENERAL INFORMATION:
APPLICANT: Steck, Peter
APPLICANT: Jassser, Samar
APPLICANT: Yung, Alfred W.K.
APPLICANT: Tavligian, Sean V.
TITLE OF INVENTION: A TUMOR SUPPRESSOR DESIGNATED TS10Q23.3
FILE REFERENCE: 2318-385
CURRENT APPLICATION NUMBER: US/10/299,003
CURRENT FILING DATE: 2002-11-19
PRIOR APPLICATION NUMBER: U.S. 09/140,749
PRIOR FILING DATE: 1998-08-26
PRIOR APPLICATION NUMBER: U.S. 08/791,115
PRIOR FILING DATE: 1997-01-30
PRIOR APPLICATION NUMBER: U.S. 60/057,750
PRIOR FILING DATE: 1997-08-26
PRIOR APPLICATION NUMBER: U.S. 60/083,563
PRIOR FILING DATE: 1998-04-30
NUMBER OF SEQ ID NOS: 64
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 52
LENGTH: 194
TYPE: PRT
ORGANISM: Mus musculus
US-10-299-003-52

Query Match
Best Local Similarity 69.8%; Score 37; DB 6; Length 194;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 QGPEVLSOP 10
DB 78 QGPEVLSKP 85

Search completed: February 16, 2003, 22:04:26
Job time : 1.40685 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:55:19 ; Search time 0.33833 Seconds
(without alignments)
2841.441 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_452

Perfect score: 53

Sequence: 1 AQPGLVLSQP 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	77.4	365	2 T02181	probable protein kinase
2	38	71.7	987	2 A75496	conserved hypothe
3	37	69.8	349	2 D83048	hypothetical prote
4	37	69.8	891	2 T22560	hypothetical prote
5	36	67.9	163	2 B82708	hypothetical prote
6	36	67.9	233	2 H72648	hypothetical prote
7	36	67.9	308	1 QOLJX1	trans-activating t
8	36	67.9	361	2 S59531	manganese peroxid
9	36	67.9	603	2 A75373	probable N-acetyl
10	36	67.9	686	2 F82871	probable FMN oxido
11	35.5	67.0	1071	2 B84062	hypothetical prote
12	35	66.0	188	1 B45540	coat protein - phy
13	35	66.0	190	2 S41553	hypothetical prote
14	35	66.0	309	2 E36596	coat protein - bel
15	35	66.0	311	2 B86939	nuclear factor I
16	35	66.0	328	2 T17789	probable isopenten
17	35	66.0	339	2 D41377	chitosanase homolo
18	35	66.0	350	2 JG7188	methanol dehydroge
19	35	66.0	357	2 PC4293	REIC protein - hum
20	35	66.0	365	2 D84848	nuclear factor 1 f
21	35	66.0	371	2 A10199	probable protein k
22	35	66.0	386	2 S43569	tRNA (5-methylamin
23	35	66.0	446	2 C69653	nuclear factor I-A
24	35	66.0	461	2 C69653	hydrolytic enzyme
25	35	66.0	498	2 S45567	nuclear factor I-A
26	35	66.0	509	2 S45565	nuclear factor I-A
27	35	66.0	509	2 JC5428	nuclear factor I f
28	35	66.0	522	2 S09996	nuclear factor I f
29	35	66.0	532	2 B36596	nuclear factor I -

30	35	66.0	587	2 JC5300	Ran GTPase activat
31	35	66.0	589	2 A36983	RNAI homolog fung1
32	35	66.0	589	2 T52070	RNAI protein homol
33	35	66.0	1075	2 G59434	KRAA0411 protein l
34	35	66.0	1108	2 T35827	probable membrane
35	35	66.0	1285	2 T14171	ataxin-2 - mouse
36	35	66.0	1445	2 T10728	Basoon protein -
37	35	66.0	3942	2 T42730	Basoon protein -
38	34	64.2	124	2 T26799	hypothetical prote
39	34	64.2	135	2 C83092	stringent starvati
40	34	64.2	152	2 T35798	hypothetical prote
41	34	64.2	188	2 P00180	Cytlb protein - Syn
42	34	64.2	236	1 A30225	regulatory protein
43	34	64.2	236	2 T51073	regulatory protein
44	34	64.2	298	2 S27965	hypothetical prote
45	34	64.2	315	2 T49982	hypothetical prote

ALIGNMENTS

RESULT 1
T02181
probable protein kinase [imported] - Arabidopsis thaliana
N:Alternate names: protein kinase homolog F14M4.11
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02181; F84910
R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.;
submitted to the EMBL Data Library, September 1998
A:Description: Arabidopsis thaliana chromosome II BAC F14M4 genomic sequence.
A:Reference number: Z14609
A:Accession: T02181
A>Status: translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-365 <R0U>
A:Cross-references: EMBL:AC004411; NID:93522932; PID:93522961
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: F84910
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-365 <STO>
A:Cross-references: GB:AE002093; NID:93522961; PID:93522961
C:Genetics:
A:Gene: F14M4.11; At2g47060
A:Map position: 2
A:Intons: 32/1; 116/3; 159/1; 250/2; 294/3; 323/3
C:Superfamily: Kinase-related transforming protein; protein kinase homology

Query Match 77.4% Score 41; DB 2; Length 365;
Best Local Similarity 100.0% Pred. No. 8.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AQPGLVLS 8
Db 166 AQPGLVLS 173

RESULT 2
A75496
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: A75496
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
M.; Shen, M.; Yamahavevan, J.J.; Lam, P.; McDonald, L.; Uitterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
A:Reference number: AF5250; MUID:20036896; PMID:10567266
A:Accession: AF5496
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-987 <MHI>
A:Cross-references: GB:AE001920; GB:AE000513; NID:96458318; PIDN:AAF10199.1; PID:9645832
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0620
A:Map position: 1

Query Match 71.7%; Score 38; DB 2; Length 987;
Best Local Similarity 70.0%; Pred. No. 80;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 AOPGVLSP 10
| | | | | : |
DB 488 AOPGVLSP 497

RESULT 3
D83048
hypothetical protein PA4774 [imported] - *Pseudomonas aeruginosa* (strain PA01)
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83048
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lam,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho
A:Reference number: AB2950; MUID:20437337; PMID:10964043
A:Accession: D83048
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <STO>
A:Cross-references: GB:AE004891; GB:AE004091; NID:995951037; PIDN:AA08160.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4774

Query Match 69.8%; Score 37; DB 2; Length 349;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AOPGVL 7
| | | | |
DB 315 AOPGVL 321

RESULT 4
T22560
hypothetical protein F53C11.5 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22560
R:Baynes, C.
submitted to the EMBL Data Library, September 1996
A:Reference number: Z19581
A:Accession: T22560
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-891 <WII>
A:Cross-references: EMBL:Z79756; PIDN:CA802122.1; GSPDB:GN00023; CESP:F53C11.5
A:Experimental source: clone F53C11
C:Genetics:
A:Gene: CESP:F53C11.5
A:Map position: 5
A:Introns: 25/3; 59/1; 137/1; 287/1; 313/1; 343/3; 420/1; 455/2; 637/2; 708/2; 831/2; 86
Query Match 69.8%; Score 37; DB 2; Length 891;

Best Local Similarity 66.7%; Pred. No. 1,1e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 OPGVLSOP 10
| | | | |
DB 49 OPGVLSOP 57

RESULT 5
B82708
hypothetical protein XP1226 [imported] - *Xylella fastidiosa* (strain 9a5c)
C:Species: *Xylella fastidiosa*
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: B82708
R:anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
A:Reference number: AB2515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: B82708
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-163 <SIM>
A:Cross-references: GB:AE003956; GB:AE003849; NID:99106196; PIDN:AAF04036.1; GSPDB:GN
A:Experimental source: strain 9a5c
R:Stimpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
B:Trons, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
as-Neto, E.; Docena, C.; El-Deiry, H.; Facincani, A.P.; Ferreira, A.D.S.
submitted to GeneBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Jungueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurama, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miranda, E.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sava
M.; Tshahko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1226

Query Match 67.9%; Score 36; DB 2; Length 163;
Best Local Similarity 87.5%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 OPGVLSOP 10
| | | | |
DB 58 OPGVLSOP 65

RESULT 6
H72648
hypothetical protein APE0622 - *Aeropyrum pernix* (strain K1)
C:Species: *Aeropyrum pernix*
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: H72648
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aero*
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: H72648
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-233 <KAN>
A:Cross-references: DBJ:AP000060; NID:95104188; PIDN:BA79592.1; PID:dt1043378; PID:9
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0622

Query Match 67.9%; Score 36; DB 2; Length 233;
Best Local Similarity 87.5%; Pred. No. 41;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AOPGPVLS 8
 | | | | |
 Db 66 ASPGPVLS 73

RESULT 7

trans-activating transcription regulator - bovine leukemia virus (fragment)
 C:Species: bovine leukemia virus, BLV
 A:Note: host Bos sp (cattle)
 C>Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 02-Jul-1998
 C:Accession: A04014
 R:Sagata, N.; Yasunaga, T.; Tsuzuku-Kawamura, J.; Ohishi, K.; Ogawa, Y.; Ikawa, Y.
 Proc. Natl. Acad. Sci. U.S.A. 82, 677-681, 1985
 A:Title: Complete nucleotide sequence of the genome of bovine leukemia virus: its evolution
 A:Reference number: A94063; MUID:85140159; PMID:2983308
 A:Accession: A04014
 A:Molecule type: DNA
 A:Residues: 1-308 <SAG>
 C:Superfamily: Leukemia virus trans-activating transcription regulator

Query Match 67.9%; Score 36; DB 1; Length 308;
 Best Local Similarity 70.0%; Pred. No. 55;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AOPGPVLS 10
 | | | | |
 Db 97 AOPGPVLS 106

RESULT 8

manganese peroxidase (EC 1.11.1.13) precursor - oyster mushroom
 C:Species: Pleurotus ostreatus (oyster mushroom)
 C>Date: 20-Jul-1996 #sequence_revision 08-Nov-1996 #text_change 11-Jun-1999
 C:Accession: S59531; S59532; S59525
 R:Asada, Y.; Watanabe, A.; Irie, T.; Nakayama, T.; Kuwahara, M.
 Biochim. Biophys. Acta 1251, 205-209, 1995
 A:Title: Structures of genomic and complementary DNAs coding for Pleurotus ostreatus man
 A:Reference number: S59525; MUID:9539442; PMID:7669812
 A:Accession: S59531
 A:Molecule type: DNA
 A:Residues: 1-361 <ASAS>
 A:Cross-references: EMBL:U21878; NID:g901983; PIDN:AAA84396.1; PID:g726132
 A:Accession: S59532
 A:Molecule type: mRNA
 A:Residues: 1-361 <ASAS>
 A:Cross-references: EMBL:U21879; NID:g732512; PIDN:AAA84397.1; PID:g732513
 A:Accession: S59525
 A:Molecule type: protein
 A:Residues: 30-59 <ASFS>
 C:Genetics:
 A:Introns: 22/1; 66/3; 86/2; 132/2; 173/1; 198/2; 210/3; 213/2; 240/2; 263/3; 267/1; 279
 C:Superfamily: ligin peroxidase
 C:Keywords: extracellular protein; glycoprotein; heme; manganese; oxidoreductase
 F:1-22/Domain: signal sequence #status predicted <SIG>
 F:23-29/Domain: propeptide #status predicted <PRO>
 F:30-361/Product: manganese (II) peroxidase #status experimental <MAT>

Query Match 67.9%; Score 36; DB 2; Length 361;
 Best Local Similarity 70.0%; Pred. No. 64;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AOPGPVLS 10
 | | | | |
 Db 347 AOPGPVLS 356

RESULT 9

A75373
 probable N-acetylmutamoyl-L-alanine amidase - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
 C:Accession: A75373
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
 M.; Shen, M.; Yamathayan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.;
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896; PMID:10567266
 A:Accession: A75373
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-603 <WHIT>
 A:Cross-references: GB:AE002007; GB:AE000513; NID:96459402; PIDN:AAE11192.1; PID:9645
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR1632
 A:Map position: 1

Query Match 67.9%; Score 36; DB 2; Length 603;
 Best Local Similarity 60.0%; Pred. No. 11e+02;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AOPGPVLS 10
 | | | | |
 Db 292 AOPGPVLS 301

RESULT 10

probable FMN oxidoreductase PA5398 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 02-Nov-2001
 C:Accession: F82971
 R:Stover, C.K.; Pham, X.Q.; Errvin, A.L.; Micoquchi, S.D.; Warren, P.; Hickey, M.J.;
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.
 .; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: F82971
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-686 <STO>
 A:Cross-references: GB:AE004952; GB:AE004091; NID:g9951718; PIDN:AGC08783.1; GSPDB:GN
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA5398
 C:Superfamily: Methylotrophus methylotrophus W3A1 trimethylamine dehydrogenase

Query Match 67.9%; Score 36; DB 2; Length 686;
 Best Local Similarity 77.8%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 OPGPVLSP 10
 | | | | |
 Db 646 OPGPVLSP 654

RESULT 11

B84062
 hypothetical protein BH3298 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: B84062
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 A:Accession: B84062
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1071 <STO>

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:54:50 ; Search time 0.192719 seconds
(Without alignments)
2152.161 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_452

Perfect score: 53
Sequence: 1 AOPGPVLSQP 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	100.0	801	T2DT_HUMAN	Q92750 homo sapien
2	37	69.8	349	DKK3_MOUSE	Q92750 mus musculu
3	37	69.8	756	IKKB_HUMAN	Q14920 homo sapien
4	37	69.8	1279	SCAP_HUMAN	Q12770 homo sapien
5	36	67.9	308	TPX1_BUYJ	P03412 bovine leuk
6	36	67.9	1942	Y054_HUMAN	P42694 homo sapien
7	35	66.0	121	SECR_HUMAN	P09663 homo sapien
8	35	66.0	188	COAT_PMYV	P36351 physalis mo
9	35	66.0	190	COAT_BMDV	P15158 belladonna
10	35	66.0	311	ISPE_MYCLE	Q96051 mycobacteri
11	35	66.0	339	MOXR_PARDE	P29901 paracoccus
12	35	66.0	350	DKK3_HUMAN	Q92750 homo sapien
13	35	66.0	446	LPID_BACSU	P39130 bacillus su
14	35	66.0	509	NEFA_HUMAN	Q12857 homo sapien
15	35	66.0	509	NEFA_HUMAN	Q12857 homo sapien
16	35	66.0	522	NEFA_CHICK	P09414 rattus norv
17	35	66.0	532	NEFA_MOUSE	P17923 gallus gall
18	35	66.0	587	RGPL_HUMAN	Q02760 mus musculu
19	35	66.0	589	RGPL_MOUSE	P46060 mus musculu
20	35	66.0	650	V411_MOUSE	Q43295 mus musculu
21	35	66.0	733	VINE_MOUSE	Q91188 mus musculu
22	34	64.2	110	GON2_SUNMU	O91686 suncus muri
23	34	64.2	236	CYS3_NEUCR	P22697 neurospora
24	34	64.2	298	Y33K_HUMAN	Q04333 homo sapien
25	34	64.2	328	ZIPA_ECO57	Q8X452 escherichia
26	34	64.2	328	ZIPA_ECOLI	P71713 escherichia
27	34	64.2	345	RLAO_PYPRAE	Q82463 pyrobaculum
28	34	64.2	360	RLPA_YERPE	Q82463 yersinia pe
29	34	64.2	838	TAC3_HUMAN	O9Y665 homo sapien
30	34	64.2	913	UBPK_HUMAN	O9Y266 homo sapien
31	33	62.3	211	SN23_HUMAN	O00161 homo sapien
32	33	62.3	228	MUTH_ECOLI	P06722 escherichia
33	33	62.3	267	XPA_XENLA	P27088 xenopus lae

34	33	62.3	313	1	CDX2_HUMAN	Q99626 homo sapien
35	33	62.3	354	1	YMA3_MYCO	Q02279 mycobacteri
36	33	62.3	464	1	CDK8_HUMAN	P49336 homo sapien
37	33	62.3	476	1	VL2_HPV32	P32607 human papil
38	33	62.3	531	1	SIS2_CANTR	Q12600 candida tro
39	33	62.3	580	1	SYN3_HUMAN	Q14994 homo sapien
40	33	62.3	603	1	BPRV_BACNO	P42779 bacteroides
41	33	62.3	626	1	PKNH_MYCTU	Q11053 mycobacteri
42	33	62.3	656	1	DNA4_STRCO	P27902 streptomyce
43	33	62.3	890	1	WFS1_HUMAN	P276024 homo sapien
44	33	62.3	890	1	WFS1_MOUSE	P56695 mus musculu
45	33	62.3	1175	1	DSRA_RAT	P55266 rattus norv

ALIGNMENTS

```

RESULT 1
ID T2DT_HUMAN STANDARD; PRT; 801 AA.
AC Q92750;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription Initiation factor TFIID 105 kDa subunit (TAIFI-105)
DE (TAIFI105) (Fragment).
GN TAF105 OR TAF2C2 OR TAF1105.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=97011146; PubMed=8658156;
RA Dikstein R., Zhou S., Tjian R.;
RT "Human TAFII 105 is a cell type-specific TFIID subunit related to
hTAFII130."
RL Cell 87:137-146(1996).
CC -!- FUNCTION: CELL TYPE-SPECIFIC SUBUNIT OF TFIID THAT MAY FUNCTION AS
A GENE-SELECTIVE COACTIVATOR IN CERTAIN CELLS. TFIID IS A
MULTIMERIC PROTEIN COMPLEX THAT PLAYS A CENTRAL ROLE IN MEDIATING
PROMOTER RESPONSES TO VARIOUS ACTIVATORS AND REPRESSORS.
CC -!- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
NUMBER OF TBP-ASSOCIATED FACTORS (TAFs). TAFII105 MAY FORM
HETERODIMERS WITH TAFII130.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: FOUND IN B CELLS BUT NOT IN OTHER TESTED
CELLS.
CC -!- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
DR EMBL; Y09321; CAAT70499.1; -
DR Genbank; HGNC:11538; TAF105.
DR MIM; 601689; -
DR InterPro; IPR003894; TAF_hom.
DR SMART; SM00549; TAFH; 1.
KW Transcription regulation; Nuclear protein.
FT NON_TER
SQ
SEQUENCE 801 AA; 85658 MW; D12B4933FEA49CD2 CRC64;
Query Match 100.0%; Score 53; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 0.059;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
|||||
1 AOPGPVLSQP 10

```

RESULT 2

ID	DRK3_MOUSE	STANDARD:	PRT:	349 AA.
AC	Q90UN9;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Dickkopf related protein-3 precursor (Dkk-3) (Dkkopf-3) (mdkx-3).			
GN	DKK3.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxId:10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99425169; PubMed=10495270;			
RA	Monaghan P.A., Kioschis P., Wu W., Zuniga A., Bock D., Pousetka A.,			
RA	Delius H., Niehrs C.;			
RT	"Dickkopf genes are co-ordinately expressed in mesodermal lineages."			
RL	Mech. Dev. 87:45-56(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20035735; PubMed=10570958;			
RA	Kirupak V.E., Sharp J.D., Jiang C., Robison K., Chickerling T.W.,			
RA	Amaravedi L., Brown D.E., Guyot D., Mays G., Leiby K., Chang B.,			
RA	Duong T., Goodenall A.D.J., Gearing D.P., Sokol S.Y., McCarthy S.A.;			
RT	"Functional and structural diversity of the human Dickkopf gene			
RL	family";			
RL	Gene 238:301-313(1999).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Liver;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Altawa K., Itawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,			
RA	Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochaya H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nakado T., Pesole G., Quackenbush J.,			
RA	Schirnl L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Botfield D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Guatinchic S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombeerts P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,			
RA	Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,			
RA	Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,			
RA	Hayashizaki Y.;			
RT	"Functional annotation of a full-length mouse cDNA collection";			
RL	Nature 409:685-690(2001).			
CC	-1- FUNCTION: INHIBITOR OF WNT SIGNALING PATHWAY (POTENTIAL).			
CC	-1- SUBCELLULAR LOCATION: Secreted			
CC	-1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN BRAIN, EYE, AND HEART.			
CC	-1- SIMILARITY: BELONGS TO THE DICKKOPF FAMILY.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/sib-ch).			
CC	or send an email to license@sib-ch).			
CC	-----			
DR	EMBL: AJ243964; CAB6011.1; -			
DR	EMBL: AF177400; AA02680.1; -			
DR	EMBL: AK004853; BAB23617.1; -			
DR	MGI:1354952; DKK3.			

FW	Developmental protein; Signal; Glycoprotein.
FT	SIGNAL
FT	CHAIN
FT	DOMAIN
FT	CARBOHYD
FT	CARBOHYD
FT	CARBOHYD
FT	CARBOHYD
FT	SEQUENCE
QY	1 A0GPEVLSOP 10
DB	32 AEPGPALNYP 41
	69.8%; Score 37; DB 1; Length 349;
	Best Local Similarity 60.0%; Pred. No. 18;
	Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
	RESULT 3
	IKKB_HUMAN
ID	IKKB_HUMAN STANDARD: PRT; 756 AA.
AC	014920; 075327;
DT	16-OCT-2001 (Rel. 40, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Inhibitor of nuclear factor kappa B kinase subunit (EC 2.7.1.-) (I-kappa-B-kinase beta) (IKBK) (IKK-beta) (IKK-B) (I-kappa-B kinase 2) (IKK2) (Nuclear factor NF-kappa-B inhibitor kinase beta) (NFKB1KB), IKKB OR IKBK.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxId	9606;
[1]	
RN	SEQUENCE FROM N.A., AND MUTAGENESIS OF LYS-44; SER-177 AND SER-181.
RP	TISSUE=Cervical Carcinoma;
RC	MEDLINE=98008813; Pubmed=9346484;
RA	Mercutio F., Zhu H., Murray B.W., Shevchenko A., Bennett B.L., Li J.W., Young D.B., Barbosa M., Mann M., Manning A., Rao A.;
RT	"IKK-1 and IKK-2: cytoskeleton-activated Ikappab kinases essential for NF-kappab activation.";
RL	Science 278:860-866(1997).
[2]	
RN	SEQUENCE FROM N.A., AND MUTAGENESIS OF LYS-44.
RP	MEDLINE=98008814; Pubmed=9346485;
RA	Woronowicz J.D., Gao X., Gao Z., Rothe M., Goeddel D.V.;
RT	"Ikappab kinase-beta: NF-kappab activation and complex formation with Ikappab kinase-alpha and NIK.";
RL	Science 278:866-869(1997).
[3]	
RN	SEQUENCE FROM N.A.
RP	TISSUE=Heart;
RC	MEDLINE=99032998; Pubmed=9813230;
RA	Hu M.C.-T., Wang Y.-P.;
RT	"Ikappab kinase-alpha and -beta genes are coexpressed in adult and embryonic tissues but localized to different human chromosomes.";
RL	Gene 222:31-40(1998).
[4]	
RN	SEQUENCE FROM N.A., AND GENE MAPPING.
RP	MEDLINE=98438415; Pubmed=9763654;
RA	Shindo M., Nakano H., Sakon S., Yagita H., Mihara M., Okumura K.;
RT	"Assignment of Ikappab kinase beta (IKBB) to human chromosome band 6p12->p11 by in situ hybridization.";
RL	Cytogenet. Cell Genet. 82:32-33(1998).
[5]	
RN	SEQUENCE OF 1-256 FROM N.A.
RP	TISSUE=Lung;
RC	Strausberg R.;
RA	Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RL	161
NP	IKK PHOSPHORYLATION.

RX MEDLINE-99038238; PubMed-9819420;
 RA Nemoto S., Didonato J.A., Lin A.;
 RT "Coordinate regulation of IkappaB kinases by mitogen-activated protein
 RT kinase kinase kinase 1 and NF-kappaB-inducing kinase.";
 RL Mol. Cell. Biol. 18:7336-7343(1998).
 RN [7]
 RN REVIEW.
 RP MEDLINE-20178139; PubMed-10712233;
 RA Jobin C., Sartor R.B.;
 RT "The I kappa B/NF-kappa B system: a key determinant of
 RT mucosal inflammation and protection.";
 RL Am. J. Physiol. 278:C451-C462(2000).
 CC -1- FUNCTION: PHOSPHORYLATES INHIBITORS OF NF-KAPPA-B THUS LEADING TO
 CC THE DISSOCIATION OF THE INHIBITOR/NF-KAPPA-B COMPLEX AND
 CC ULTIMATELY THE DEGRADATION OF THE INHIBITOR.
 CC -1- SUBUNIT: PREFERENTIALLY FOUND AS A HETERODIMER WITH IKK-ALPHA BUT
 CC ALSO AS A HOMODIMER. DIRECTLY INTERACTS WITH IKK-GAMMA/NEMO.
 CC HETERODIMERS FORM THE ACTIVE COMPLEX. THE TRIPARTITE COMPLEX CAN
 CC ALSO BIND TO MEK1, MAPK14/NIK, IKAP AND IKB-ALPHA-P65-P50
 CC COMPLEX. PHOSPHORYLATED IKB-ALPHA IS FURTHER RELEASED FROM THE
 CC COMPLEX.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART, PLACENTA, SKELETAL
 CC MUSCLE, KIDNEY, PANCREAS, SPLEEN, THYMOS, PROSTATE, TESTIS AND
 CC PERIPHERAL BLOOD.
 CC -1- PFM: PHOSPHORYLATED BY MEK1 AND PROBABLY ALSO BY MAPK14/NIK.
 CC WEAKLY AUTOPHOSPHORYLATED.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC IKAPPAB KINASE SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF029684; AAC51860.1; -;
 DR EMBL: AF080158; AAD08997.1; -;
 DR EMBL: AF031416; AAC64675.1; -;
 DR EMBL: BC006231; AAH06231.1; -;
 DR HSSP: Q63450; IA06.
 DR Genew; HGNC:5960; IKBK.
 DR MIM: 603258; -;
 DR InterPro: IPR000719; Euk.pkinase.
 DR InterPro: IPR002290; Ser-thr.pkinase.
 DR InterPro: IPR001245; Tyr.pkinase.
 DR Pfam: PF00069; pkinase.1.
 DR Pfam: PF00240; ubiqtulin.1.
 DR PRINTS: PRO0109; TYRKINASE.
 DR ProDom: PD000001; Euk.pkinase.1.
 DR ProSite: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST.1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM.1.
 DR Transferrase; Serine/threonine-protein kinase; ATP-binding;
 DR phosphorylation.
 KW DOMAIN
 FT 15 300 PROTEIN KINASE.
 FT DOMAIN 458 479 LEUCINE-ZIPPER (POTENTIAL).
 FT 737 742
 FT NP_BIND 21 29 ATP (BY SIMILARITY).
 FT BINDING 44 44 ATP (BY SIMILARITY).
 FT ACT_SITE 145 145 BY SIMILARITY.
 FT MOD_RES 23 23 PHOSPHORYLATION (BY SIMILARITY).
 FT MOD_RES 177 177 PHOSPHORYLATION.
 FT MOD_RES 181 181 PHOSPHORYLATION.
 FT MUTAGEN 44 44 K->A: LOSS OF KINASE ACTIVITY AND NO
 FT MUTAGEN 177 177 EFFECT ON BINDING TO NIK.
 FT MUTAGEN 177 177 S->A: DECREASE OF ACTIVITY.
 FT MUTAGEN 181 181 S->E: FULL ACTIVATION.
 FT MUTAGEN 181 181 S->A: DECREASE OF ACTIVITY.
 FT MUTAGEN 181 181 S->E: FULL ACTIVATION.
 FT CONFLICT 231 255 WSKRSEVDIVSEDLNCTVKE -> CVRMMPGVVAHS

FT CONFLICT 425 425 CNPSTLGGRGHWI (IN REF. 5).
 FT SEQUENCE 756 AA; 86563 MW; F9CADF671AE9E1AE CRC64;
 SQ
 Query Match 69.8%; Score 37; DB 1; Length 756;
 Best Local Similarity 60.0%; Pred. No. 41;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 AOPGPVLSQP 10
 Db 682 SQPGQLMSQP 691
 RESULT 4
 SCAP_HUMAN STANDARD; PRT; 1279 AA.
 ID SCAP_HUMAN
 AC Q12770;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sterol regulatory element binding protein cleavage-activating protein
 DE (SREBP cleavage-activating protein) (SCAP).
 DE SCAP OR KIAA0199.
 GN SCAP
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 [1]
 [1] SEQUENCE FROM N.A.
 RP TISSUE-Hepatoma;
 RX MEDLINE-20037503; PubMed-10570913;
 RA Nakajima T., Hamakubo T., Kodama T., Inazawa J., Eml M.;
 RT "Genomic structure and chromosomal mapping of the human sterol
 RT regulatory element binding protein (SREBP) cleavage-activating
 RT protein (SCAP) gene";
 RL J. Hum. Genet. 44:402-407(1999).
 RN [2]
 [2] SEQUENCE OF 3-1279 FROM N.A.
 RP TISSUE-Bone marrow;
 RX MEDLINE-96281124; PubMed-8724849;
 RA Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
 RT "Prediction of the coding sequences of unidentified human genes. V.
 RT The coding sequences of 40 new genes (K1AA0161-K1AA0200) deduced by
 RT analysis of cDNA clones from human cell line KG-1.";
 RL DNA Res. 3:17-24(1996).
 RN [3]
 [3] SEQUENCE OF 346-1279 FROM N.A.
 RP TISSUE-Teratocarcinoma;
 RX Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Magatsuma M., Hosokari T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
 RA Minomiya K., Iwayanagi T.;
 RA "NEDO human cDNA sequencing project.";
 RT Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
 RL [4]
 [4] VARIANT ILE-798.
 RP MEDLINE-20037509; PubMed-10570919;
 RX Iwaki K., Nakajima T., Ota N., Eml M.;
 RA "A common ile798val polymorphism of the human SREBP cleavage-
 RT activating protein (SCAP) gene.";
 RL J. Hum. Genet. 44:421-422(1999).
 CC -1- FUNCTION: Sterol sensor. Necessary for the proteolytic activation
 CC of SREBPs by site-1 protease in the Golgi.
 CC -1- SUBUNIT: In a tight complex with SREBP (By similarity).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Moves from the
 CC endoplasmic reticulum to the Golgi in the absence of sterols (By
 CC similarity).
 CC -1- SIMILARITY: CONTAINS 1 STEROL SENSOR DOMAIN (SSD).
 CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D83782; BAA12111.1; -
 DR EMBL: AK027402; BAB55088.1; ALT_INIT.
 DR MIM: 601510; -
 DR InterPro: IPR000731; HMGCR/Patch_5TM.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 12.
 DR PRINTS: PR00320; GPROTEINBRPT.
 DR SMART: SM00320; WD40; 10.
 DR PROSITE: PS00156; SSD; 1.
 DR PROSITE: PS50082; WD_REPEATS_2; 1.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 DR PROSITE: PS00678; WD_REPEATS_1; 1.
 KW Lipid metabolism; Cholesterol metabolism; Transmembrane;
 KW Endoplasmic reticulum; Golgi stack; Repeat; WD repeat; Glycoprotein;
 KW Polymorphism.
 FT DOMAIN 1 18
 FT TRANSMEM 19 39
 FT DOMAIN 40 279
 FT TRANSMEM 280 300
 FT DOMAIN 301 312
 FT TRANSMEM 313 333
 FT DOMAIN 334 344
 FT TRANSMEM 345 365
 FT DOMAIN 366 401
 FT TRANSMEM 402 422
 FT DOMAIN 423 423
 FT TRANSMEM 424 444
 FT DOMAIN 445 518
 FT TRANSMEM 519 539
 FT DOMAIN 540 709
 FT TRANSMEM 710 730
 FT DOMAIN 731 1279
 FT TRANSMEM 284 442
 FT REPEAT 771 811
 FT REPEAT 952 1002
 FT REPEAT 1005 1042
 FT REPEAT 1077 1114
 FT REPEAT 1117 1155
 FT REPEAT 1158 1195
 FT REPEAT 1197 1235
 FT DOMAIN 722 729
 FT DOMAIN 747 750
 FT DOMAIN 861 864
 FT CARBOHYD 263 263
 FT CARBOHYD 590 590
 FT CARBOHYD 641 641
 FT VARIANT 798 798
 FT CONFLICT 617 617
 FT CONFLICT 753 753
 FT CONFLICT 817 817
 FT CONFLICT 941 941
 FT CONFLICT 1035 1035
 FT CONFLICT 1080 1080
 FT SEQUENCE 1279 AA; 139728 MW; 4CD73543B7D2ACB4 CRC64;
 Query Match 69.8%; Score 37; DB 1; Length 1279;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 PGVLSQ 9
 DB 938 PGVLSQ 944

RESULT 5
 YPXL_BLVU STANDARD; PRT; 308 AA.
 ID YPXL_BLVU
 AC P03412;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-MAY-1992 (Rel. 22, Last annotation update)
 DE Hypothetical PXLV-I protein (fragment).
 OS Bovine leukemia virus (Japanese isolate BLV-1) (BLV).
 OC Viruses; Retroviridae; Retroviridae; Deltaretrovirus.
 NCBI_Taxid=11907;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=65140159; PubMed=2983308;
 RA Segata N., Yasunaga T., Tsuzuku-Kawamura J., Ohishi K., Ogawa Y.,
 RA Ikawa Y.;
 RT "Complete nucleotide sequence of the genome of bovine leukemia virus:
 RT its evolutionary relationship to other retroviruses.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:677-681(1985).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: K02120; -; NOT_ANNOTATED_CDS.
 DR PIR: A04014; GOLLX1.
 KW Hypothetical protein.
 FT NON_TER 1
 FT SEQUENCE 308 AA; 34191 MW; 082AD897E614672B CRC64;
 Query Match 67.9%; Score 36; DB 1; Length 308;
 Best Local Similarity 70.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ACPGVLSQ 10
 DB 97 ACPGVLSQ 106

RESULT 6
 Y054_HUMAN STANDARD; PRT; 1942 AA.
 ID Y054_HUMAN
 AC P42694;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein KIAA0054.
 DE KIAA0054.
 GN KIAA0054.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE=96051398; PubMed=7584044;
 RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
 RA Seki N., Kawarabayashi Y., Ishikawa K.-I., Tabata S.;
 RT "Prediction of the coding sequences of unidentified human genes. II.
 RT The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by
 RT analysis of cDNA clones from human cell line KG-1.";
 RL DNA Res. 1:223-229(1994).
 CC -----
 CC 1- SIMILARITY: CONTAINS 1 C3H1-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: D29677; BAA06147.1; -

DR Genew: HGNC:16878; HELZ.

DR InterPro: IPR000571; Znf_CCH.

DR Pfam: PF00642; Zf_CCH.1.

DR SMART: SM00356; Znf_C3H1.1.

KW Hypothetical protein; zinc-finger.

SO SEQUENCE 1942 AA; 218970 MW; 94D13601A02335E0 CRC64;

Query Match

Best Local Similarity 67.9%; Score 36; DB 1; Length 1942;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AOPGPVL 7

Db 1469 AOPGPVL 1475

RESULT 7

SECR_HUMAN STANDARD; PRT; 121 AA.

AC P09683;

DT 01-MAR-1989 (Rel. 10, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Secretin precursor.

GN SCT.

OS Homo sapiens (human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20515579; PubMed=11060443;

RA Wiltmore T.E., Holloway J.L., Lofton-Day C.E., Maurer M.F., Chen L., Quinton T.J., Vincent J.B., Scherer S.W., Lok S.;

RT "Human secretin (SCT): gene structure, chromosome location, and distribution of mRNA".

RT Cytogenet. Cell Genet. 90:47-52(2000).

RL [2]

RN SEQUENCE OF 28-54.

RA Carlquist M., Joernvall H., Forssmann W.-G., Thulin L., Johansson C., Mutt V.;

RT "Human secretin is not identical to the porcine/bovine hormone.";

RL ICS Med. Sci. 13:217-218(1985).

CC -1- FUNCTION: STIMULATES FORMATION OF NAHCO(3)-RICH PANCREATIC JUICE AND SECRETION OF NAHCO(3)-RICH BILE AND INHIBITS HCL PRODUCTION BY THE STOMACH.

CC -1- SIMILARITY: BELONGS TO THE GLUCAGON FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AF244355; AAC31443.1; -

DR PIR: S07443; S07443.

DR Genew: HGNC:10607; SCT.

DR MIM: 182099; -

DR InterPro: IPR000532; Glucagon.

DR Pfam: PF00123; hormone2; 1.

DR SMART: SM0070; GLUCA.1.

DR PROSITE: PS00260; GLUCAGON.1.

KW Glucagon family; Hormone; Amidation; Cleavage on pair of basic residues; Signal.

FT SIGNAL 1

FT PEPTIDE 18

FT MOD_RES 34 54 SECRETIN.

FT MOD_RES 34 54 AMIDATION (G-55 PROVIDE AMIDE GROUP).

SO SEQUENCE 121 AA; 13016 MW; 44BDB4EFC0E161CF CRC64;

Query Match

Best Local Similarity 66.0%; Score 35; DB 1; Length 121;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GPGVLSQP 10

Db 102 GPGVLSQP 109

RESULT 8

COAT_PHV STANDARD; PRT; 188 AA.

ID COAT_PHV

AC P36351;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-JUN-1994 (Rel. 29, Last annotation update)

DE Coat protein (Virion protein).

OS Physalis mottle virus (PhMV) (Belladonna mottle virus-Iowa).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tymoviruses.

OX NCBI_TaxID=72539;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=9222369; PubMed=1562236;

RA Jacob A.N.K., Murthy M.R., Savithri H.S.;

RT "Nucleotide sequence of the 3' terminal region of belladonna mottle virus-Iowa (renamed Physalis mottle virus) RNA and an analysis of the relationships of tymoviral coat proteins.";

RL Arch. Virol. 123:367-377(1992).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=93212529; PubMed=8460497;

RA Kekuda R., Karande A.A., Jacob A.N.K., Savithri H.S.;

RT "Architecture of physalis mottle tymovirus as probed by monoclonal antibodies and cross-linking studies.";

RL Virology 193:959-966(1993).

CC -1- SUBUNIT: THE VIRUS COAT IS COMPOSED OF 180 COPIES OF THE COAT PROTEIN ARRANGED IN A ICOSAHERAL SHELL.

CC -1- SIMILARITY: TO COAT PROTEIN FROM OTHER TYMOVIRUSES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: S97776; AAB21997.1; -

DR PIR: B45540; B45540.

DR HSSP: P03608; IADV.

DR InterPro: IPR000574; Tymo_coat.

DR Pfam: PF00983; Tymo_coat; 1.

KW Coat protein.

SO SEQUENCE 188 AA; 19974 MW; DIDC64831FE5DC1 CRC64;

Query Match

Best Local Similarity 66.0%; Score 35; DB 1; Length 188;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GPGVLSQP 10

Db 17 PGSILSOP 24

RESULT 9

COAT_BMDV STANDARD; PRT; 190 AA.

ID COAT_BMDV

AC P15158;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Coat protein (Viron protein).
 OS Belladonna mottle virus (BMV).
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Tymovirus.
 CC NCBI_TaxID=12149;
 CC [1]
 CC [1]
 CC MEDLINE=89197927; PubMed=2467911;
 CC Suryanarayana S., Rao N.A., Murthy M.R.N., Savithri H.S.;
 CC Primary structure of belladonna mottle virus coat protein.";
 CC J. Biol. Chem. 264:6273-6279(1989).
 CC [2]
 CC SEQUENCE OF 144-186.
 CC Suryanarayana S., Rao D.R., Rao N.A., Savithri H.S.;
 CC Partial amino acid sequence of belladonna mottle virus coat
 CC protein.";
 CC Biochem. Int. 15:65-72(1987).
 CC -1- SUBUNIT: THE VIRUS COAT IS COMPOSED OF 180 COPIES OF THE COAT
 CC PROTEIN ARRANGED IN A ICOSAEDRAL SHELL.
 CC -1- SIMILARITY: TO COAT PROTEIN FROM OTHER TYMOVIRUSES.
 CC PIR: S02207; S02207.
 CC HSP: P03608; IAUY.
 CC InterPro: IPR000574; Tymo_coat.
 CC Pfam: PF00983; Tymo_coat; 1.
 CC Coat protein; Acetylation.
 CC MOD_RES 1 1 R -> K (IN REF. 2).
 CC CONFLICT 179 179 KPM -> SFQ (IN REF. 2).
 CC CONFLICT 184 186
 CC SEQUENCE 190 AA; 20215 MW; 20EE23D2D83E14 CRC64;
 SQ
 Query Match 66.0%; Score 35; DB 1; Length 190;
 Best Local Similarity 75.0%; Pred. No. 22;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 3 PGVUSQP 10
 Db 17 PGSIISQP 24
 RESULT 10
 ID ISPE_MYCLE STANDARD; PRT; 311 AA.
 AC Q9CD51;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)
 DE (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).
 GN ISPE OR M0242.
 OS Mycobacterium leprae.
 CC Bacteria; Actinobacteria; Actinobacteridae; Actinobacteriales;
 CC Actinomycetales; Corynebacteriaceae; Mycobacterium.
 CC NCBI_TaxID=1769;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN-TN;
 CC MEDLINE=21128732; PubMed=11234002;
 CC Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
 CC Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 CC Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 CC Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 CC Holtroyd S., Hornby T., Jagsels K., Lacroix C., Maclean J., Moule S.,
 CC Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 CC Rutter S., Seeger K., Simon S., Simmonds M., Skellon J., Squares R.,
 CC Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 CC Barrell B.G.;
 CC "Massive gene decay in the leprosy bacillus.";
 CC Nature 409:1007-1011(2001).
 CC -1- FUNCTION: Catalyzes the phosphorylation of the position 2 hydroxy
 CC group of 4-diphosphocytidylyl-2C-methyl-D-erythritol (By
 CC similarity).
 CC -1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
 CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-
 CC methyl-D-erythritol.

CC -1- PATHWAY: Nomenclature terpenoid biosynthesis pathway; fourth
 CC step.
 CC -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC CC
 CC EMBL: AL583917; CAC29750.1; -
 CC Leprema; M0242;
 CC InterPro: IPR001745; GMPkinase_ATP.
 CC InterPro: IPR004424; ISPE.
 CC Pfam: PF00288; GMP_kinases; 1.
 CC TIGRPFAM: TIGR00154; ISPE; 1.
 CC Transferrase; Kinase; Isoprene biosynthesis; ATP-binding;
 CC Complete proteome.
 CC NP_BIND 98 108 ATP (POTENTIAL).
 CC SEQUENCE 311 AA; 31793 MW; DF9F572BAC81C22 CRC64;
 SQ
 Query Match 66.0%; Score 35; DB 1; Length 311;
 Best Local Similarity 75.0%; Pred. No. 36;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AQPGVUS 8
 Db 201 AERGVILA 208
 RESULT 11
 ID MOXR_PARDE STANDARD; PRT; 339 AA.
 AC P29901;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-APR-1993 (Rel. 25, Last annotation update)
 DE Moxr protein.
 GN MOXR.
 OS Paracoccus denitrificans.
 CC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 CC Paracoccus.
 CC NCBI_TaxID=266;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC MEDLINE=92041581; PubMed=1657871;
 CC van Spanning R.J.M., Wansell C.W., de Boer T., Hazelaar M.J.,
 CC Anazawa H., Harms N., Oltmann L.F., Stouthamer A.H.;
 CC "Isolation and characterization of the moxr, moxg, moxi, and moxr
 CC genes of Paracoccus denitrificans: Inactivation of moxr, moxg, and
 CC moxr and the resultant effect on methylotrophic growth.";
 CC J. Bacteriol. 173:6948-6961(1991)
 CC -1- FUNCTION: MAY BE INVOLVED IN THE REGULATION OF FORMATION OF ACTIVE
 CC METHANOL DEHYDROGENASE.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: WITH THE CORRESPONDING PROTEIN IN M.EXTORQUENS.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC CC
 CC EMBL: M57684; -; NOT_ANNOTATED_CDS.
 CC PIR: DA1377; DA1377.
 CC InterPro: IPR003442; UPF0079.
 CC Pfam: PF02367; UPF0079; 1.
 CC Methanol utilization; ATP-binding.
 CC NP_BIND 47 54 ATP (POTENTIAL).

SQ SEQUENCE 339 AA; 36923 MW; 9ACFB52B6FE5FC64 CRC64;
 Query Match 66.0%; Score 35; DB 1; Length 339;
 Best Local Similarity 75.0%; Pred. No. 40;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 QPGPYLSQP 9
 Db 97 EPGPYLQ 104
 RESULT 12
 DRK3_HUMAN STANDARD; PRT; 350 AA.
 AC Q9UBP4; Q9UBP7;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Dickkopf related protein-3 precursor (Dkk-3) (Dkkopf-3) (hDkk-3).
 GN DRK3 OR REIC.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 ON NCBI_TaxID=9606;
 RX MEDLINE=20035735; PubMed=10570958;
 RA Krupnik V.E., Sharp J.D., Jiang C., Robison K., Chickering T.W.,
 RA Amaravadi L., Brown D.E., Guyot D., Mays G., Leiby K., Chang B.,
 RA Duong T., Goodearl A.D.J., Gearling D.P., Sokol S.Y., McCarthy S.A.;
 RT "Functional and structural diversity of the human Dickkopf gene
 RT family.";
 RT Gene 238:301-313(1999).
 RL [2]
 RP SEQUENCE FROM N.A.
 RA Tanaka S., Sugimachi K.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20119095; PubMed=10652205;
 RA Tsuji T., Miyazaki M., Sakaguchi M., Inoue Y., Namba M.;
 RT "A REIC gene shows down-regulation in human immortalized cells and
 RT human tumor-derived cell lines.";
 RL Biochem. Biophys. Res. Commun. 268:20-24(2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Tate G., Mitsuya T.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RX PubMed=11814687;
 RA Kobayashi K., Ouchida M., Tsuji T., Hanafusa H., Miyazaki M.,
 RA Namba M., Shimizu N., Shimizu K.;
 RT "Reduced expression of the REIC/Dkk-3 gene by promoter-
 RT hypermethylation in human tumor cells.";
 RL Gene 282:151-158(2002).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RA Strausberg R.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INHIBITOR OF WNT SIGNALING PATHWAY (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN HEART, BRAIN, AND SPINAL
 CC CORD.
 CC -1- PTM: N-GLYCOSYLATED.
 CC -1- SIMILARITY: BELONGS TO THE DICKKOPF FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF177396; BAA02676.1; -;
 DR EMBL; AB033421; BAA85468.1; -;
 DR EMBL; AB034203; BAA90548.1; -;
 DR EMBL; AB035182; BAA87044.2; -;
 DR EMBL; AB045205; BAA87044.2; JOINED.
 DR EMBL; AB045206; BAA87044.2; JOINED.
 DR EMBL; AB045207; BAA87044.2; JOINED.
 DR EMBL; AB045208; BAA87044.2; JOINED.
 DR EMBL; AB045209; BAA87044.2; JOINED.
 DR EMBL; AB045210; BAA87044.2; JOINED.
 DR EMBL; AB057591; BAA84360.1; -;
 DR EMBL; AB057804; BAA84361.1; -;
 DR EMBL; BC007660; BAA87660.1; -;
 DR Genew; HGNC:2893; DKK3.
 DR MIM; 605416; -;
 KW Developmental protein; Signal; Glycoprotein.
 FT SIGNAL 1 16 POTENTIAL.
 FT CHAIN 17 350 DICKKOPF RELATED PROTEIN-3.
 FT DOMAIN 147 195 DKK-TYPE CYS-1.
 FT DOMAIN 208 284 DKK-TYPE CYS-2.
 FT DOMAIN 338 343 POLY-ALA.
 FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 106 106 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 204 204 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 335 335 G->R (IN REF. 4).
 SQ SEQUENCE 350 AA; 38291 MW; 72F504122B40AFEE CRC64;
 QY 2 QPGPYLSQP 10
 Db 33 KPGPALSYR 41
 RESULT 13
 LPLD_BACSU STANDARD; PRT; 446 AA.
 ID LPLD_BACSU
 AC P39130; Q31532;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE LplD protein.
 GN LPLD.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 ON NCBI_TaxID=1423;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni C.,
 RA Azevedo V., Berrtero M.G., Bessieres P., Biotolin A., Borchert S.,
 RA Bottliss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,

RA Joris B., Karamata D., Kasahara Y., Kjaer-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
 RA Paro V., Pohl T.M., Portetle D., Porwollik S., Prescott A.M.,
 RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scollone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpsita S., Tognoni K.,
 RA Taseko V., Uchiyama S., Vandenpol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambolt R., Wedler E., Wedler H., Weitzenger T.,
 RA Winters P., Wipac A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa K., Zumbstein E.,
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*.";
 RT Nature 390:249-256(1997).
 CC -1- COFACTOR: NAD AND A DIVALENT METAL ION (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO FAMILY 4 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: L19165; AAA22577.1; -
 DR EMBL: 299107; CAB12532.1; -
 DR Subtilast; BG10858; lpid.
 DR InterPro: IPR001088; GH_4.
 DR Pfam: PF02056; Glyco_hydro_4; 1.
 DR PRINTS: PR00732; GLHYDRLASE4.
 DR ProDom: PD006892; GH_4; 1.
 DR PROSITE: PS01324; GLYCOSYL_HYDROL_F4; 1.
 DR Hydrolase; Glycosidase; NAD; Complete proteome.
 KM NP_BIND 12 44
 FT ACT_SITE 44 44
 FT ACT_SITE 113 113
 FT ACT_SITE 367 367
 FT CONFLICT 66 66
 FT CONFLICT 154 154
 FT SEQUENCE 446 AA; 49486 MW; 691FECC371B9F5B CRC64;
 SQ
 Query Match 66.0%; Score 35; DB 1; Length 446;
 Best Local Similarity 75.0%; Pred. No. 53;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 3 PGPVLSOP 10
 DB 282 PGPVLSOP 289
 RESULT 14
 NFIA_HUMAN STANDARD; PRT; 509 AA.
 AC Q12857; Q9P2A9; Q9H3X9;
 DT 16-OCT-2001 (Rel. 40; Created)
 DT 16-OCT-2001 (Rel. 40; Last sequence update)
 DT 16-OCT-2001 (Rel. 40; Last annotation update)
 DE Nuclear factor 1 A-type (Nuclear factor 1/A) (NF1-A) (NF1-A) (NF1-A)
 DE (CCAAT-box binding transcription factor) (CTF) (TGCCA-binding
 DE protein).
 GN NFIA OR KIAA1439.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]

RP SEQUENCE FROM N.A.
 RC TISSUE-Brain;
 RX MEDLINE=20181126; PubMed=10718198;
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
 RT Prediction of the coding sequences of unidentified human genes. XVI.
 RT The complete sequences of 150 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 7:65-73(2000).
 RN [2]
 RP SEQUENCE OF 19-243 FROM N.A.
 RX MEDLINE=96070434; PubMed=7590749;
 RA Qian F., Kruse U., Lichter P., Stipel A.E.;
 RT Chromosomal localization of the four genes (NF1, B, C, and X) for
 RT the human transcription factor nuclear factor 1 by FISH.";
 RN Genomics 28:66-73(1995).
 RP SEQUENCE OF 235-509 FROM N.A.
 RA Donnelly S.;
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: RECOGNIZES AND BINDS THE PALINDROMIC SEQUENCE 5'-
 CC TTGGGNNNNNGCCAA-3' PRESENT IN VIRAL AND CELLULAR PROMOTERS AND IN
 CC THE ORIGIN OF REPLICATION OF ADENOVIRUS TYPE 2. THESE PROTEINS ARE
 CC INDIVIDUALLY CAPABLE OF ACTIVATING TRANSCRIPTION AND REPLICATION.
 CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE CTF/NF-1 FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 DWA/MH1 DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AB037860; BAA92677.1; ALF_INIT.
 DR EMBL: U07809; AAA93124.1; -
 DR EMBL: AL096888; CAC10291.1; -
 DR TRANSFAC: T04933; -
 DR GeneW: HGNC:7784; NF1A.
 DR MIM: 600727; -
 DR InterPro: IPR000647; CTF_NF1_family.
 DR InterPro: IPR003619; DwaRfln_A.
 DR Pfam: PF00859; CTF_NF1; 1.
 DR SMART: SM00523; DWA; 1.
 DR PROSITE: PS00349; CTF_NF1; 1.
 DR Transcription regulation; DNA replication; DNA-binding; Activator;
 KM Nuclear protein; Multigene family.
 FT DOMAIN 67 175
 FT CONFLICT 186 186
 FT CONFLICT 240 243
 FT SEQUENCE 509 AA; 55944 MW; 42090C6B829F87 CRC64;
 SQ
 Query Match 66.0%; Score 35; DB 1; Length 509;
 Best Local Similarity 66.7%; Pred. No. 62;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 QGPVLSOP 10
 DB 378 QGPVLSOP 386
 RESULT 15
 NFIA_RAT STANDARD; PRT; 509 AA.
 AC P09414; O63782;
 DT 01-MAR-1989 (Rel. 10; Created)
 DT 01-NOV-1997 (Rel. 35; Last sequence update)
 DT 16-OCT-2001 (Rel. 40; Last annotation update)
 DE Nuclear factor 1 A-type (Nuclear factor 1/A) (NF1-A) (NF1-A) (NF1-A)
 DE (CCAAT-box binding transcription factor) (CTF) (TGCCA-binding
 DE protein).
 GN NFIA
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]

Search completed: February 16, 2003, 21:56:23
 Job time : 2.19272 secs

GN NFIA OR NF1L21.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Sprague-Dawley; TISSUE-Liver;
 RX MEDLINE=96429753; PubMed=8832903;
 RA Monaci P., Nuzzo M., Stampfl S., Tollervey D., de Simone V.,
 RA Nicotia A.;
 RT "A complex interplay of positive and negative elements is responsible
 for the different transcriptional activity of liver NF1 variants.";
 RL Mol. Biol. Rep. 21:147-158(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Sprague-Dawley; TISSUE-Liver;
 RX MEDLINE=97244636; PubMed=9089412;
 RA Osada S., Daimon S., Ikeda T., Nishihara T., Yano K., Yamasaki M.,
 RA Imagawa M.;
 RT "Nuclear factor 1 family proteins bind to the silencer element in the
 rat glutathione transferase P gene.";
 RL J. Biochem. 121:355-363(1997).
 RN [3]
 RP SEQUENCE OF 5-509 FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-Liver;
 RX MEDLINE=89030607; PubMed=3053160;
 RA Paonessa G., Gonnari F., Frank R., Cortese R.;
 RT "Purification of a NF1-like DNA-binding protein from rat liver and
 cloning of the corresponding cDNA.";
 RL EMBO J. 7:3115-3123(1988).
 CC -1- FUNCTION: RECOGNIZES AND BINDS THE PALINDROMIC SEQUENCE 5'-
 TTGGCANNNGCCAA-3' PRESENT IN VIRAL AND CELLULAR PROMOTERS AND IN
 THE ORIGIN OF REPLICATION OF ADENOVIRUS TYPE 2. THESE PROTEINS ARE
 INDIVIDUALLY CAPABLE OF ACTIVATING TRANSCRIPTION AND REPLICATION.
 CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE CTF/NF-1 FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 DNA/MH1 DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X84209; CAAS8995.1; -;
 DR EMBL; D78017; BAAL1203.1; -;
 DR EMBL; X13167; CA31565.1; -;
 DR PIR; S01300; S01300.
 DR TRANSFAC; T00535; -;
 DR TRANSFAC; T00539; -;
 DR TRANSFAC; T02296; -;
 DR InterPro; IPR000647; CTF_NF1_family.
 DR InterPro; IPR003619; Dwarfin_A.
 DR Pfam; PF00859; CTF_NF1; 1.
 DR SMART; SM00523; DNA; 1.
 DR PROSITE; PS00349; CTF_NF1; 1.
 KW Transcription regulation; DNA replication; DNA-binding; Activator;
 KW Nuclear protein; Multigene family.
 FT DOMAIN 67 175 DNA.
 SO SEQUENCE 509 AA; 55976 MW; 59120C7090229F87 CRC64;

Query Match 66.0%; Score 35; DB 1; Length 509;
 Best Local Similarity 66.7%; Pred. No. 62;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 OPGPVLSP 10
 DB 378 OPGPFSHP 386

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:31:01 ; Search time 0.702355 Seconds
(without alignments)
2933.659 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_452

Perfect score: 53
Sequence: 1 AQPcPVLISQP 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriapi:*
17: sp_archaeapi:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	77.4	354	10	041328 lycopersico
2	41	77.4	360	10	091KY4 glycine max
3	41	77.4	360	10	091KY3 glycine max
4	41	77.4	364	10	091LUT0 arabidopsis
5	41	77.4	365	10	080719 arabidopsis
6	39	73.6	258	5	090191 leishmania
7	39	73.6	797	4	09UGS3 homo sapien
8	38	71.7	244	4	08WYV2 homo sapien
9	38	71.7	265	2	093G77 streptomyce
10	38	71.7	397	4	096JW9 homo sapien
11	38	71.7	397	4	096JW9 homo sapien
12	38	71.7	583	16	08XUK6 ralsstonia s
13	38	71.7	583	16	08XUK6 ralsstonia s
14	38	71.7	649	11	0991L9 mus musculu
15	38	71.7	728	4	098T94 homo sapien
16	38	71.7	818	4	09C0B0 homo sapien

17	38	71.7	824	4	08TF48	08TF48 homo sapien
18	38	71.7	942	4	096K70	096K70 homo sapien
19	38	71.7	987	16	09RWP6	09RWP6 detriococcus
20	37	69.8	349	11	09CZ31	09CZ31 mus musculu
21	37	69.8	349	16	09HV34	09HV34 pseudomonas
22	37	69.8	371	10	09AMP9	09AMP9 oryza sativ
23	37	69.8	465	5	08WS46	08WS46 okopileura
24	37	69.8	553	6	095KH7	095KH7 macaca fasc
25	37	69.8	595	10	08SLW9	08SLW9 oryza sativ
26	37	69.8	891	5	093763	093763 caenorhabdl
27	37	69.8	905	4	08WU41	08WU41 homo sapien
28	36	67.9	100	6	077624	077624 bos taurus
29	36	67.9	163	16	09PE02	09PE02 xylella fas
30	36	67.9	180	16	08XRY8	08XRY8 ralsstonia s
31	36	67.9	233	17	09YEF4	09YEF4 aeropyrum p
32	36	67.9	309	15	09IG01	09IG01 bovine leuk
33	36	67.9	309	15	085493	085493 bovine leuk
34	36	67.9	309	15	056232	056232 pleurotus o
35	36	67.9	361	3	012332	012332 pleurotus o
36	36	67.9	361	3	094753	094753 pleurotus e
37	36	67.9	362	13	092074	092074 gallus gall
38	36	67.9	362	13	092074	092074 gallus gall
39	36	67.9	374	4	09NS36	09NS36 homo sapien
40	36	67.9	511	16	09JRY1	09JRY1 streptomyce
41	36	67.9	580	2	09L705	09L705 kyococcus
42	36	67.9	592	11	097301	097301 mus musculu
43	36	67.9	603	16	09RTX0	09RTX0 detriococcus
44	36	67.9	686	16	09HMG6	09HMG6 pseudomonas
45	36	67.9	867	4	09NU50	09NU50 homo sapien

ALIGNMENTS

RESULT 1
ID 041328 PRELIMINARY; PRT; 354 AA.
AC 041328;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Pto Kinase Interactor 1.
GN P11.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO GRANDE-PTOR;
RX MEDLINE=96107331; PubMed=8521516;
RA Zhou J., Loh Y.T., Bressan R.A., Martin G.B.;
RT "The tomato gene P11 encodes a serine/threonine kinase that is
RT phosphorylated by Pto and is involved in the hypersensitive
RT response.";
RL Cell 83:925-935(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO GRANDE-PTOR;
RA Martin G.B.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL: U28007; AAC61805.1; -;
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR002203; Intein.
DR InterPro: IPR002290; Ser_thr_kinase.
DR InterPro: IPR004040; STY_Pkinase.
DR Pfam: PF00069; pkinase; 1.
DR ProDom: PD000001; Euk_Pkinase; 1.
DR SMART: SM00221; STYK; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE: PSS0011; PROTEIN_KINASE_DOM; 1.

DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS00881; PROTEIN_SPLICING; UNKNOWN_1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 354 AA; 38540 MW; A6C5346175FCE97 CRC64;
Query Match 77.4%; Score 41; DB 10; Length 354;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AOPGPVLS 8
Db 161 AOPGPVLS 168
|||||
RESULT 2
ID 09LKY4 PRELIMINARY; PRT; 360 AA.
AC 09LKY4;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE PtlI kinase-like protein.
GN PtlIA.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21036115; PubMed=11185558;
RA Staswick P.;
RT "Two expressed soybean genes with high sequence identity to tomato
PtlI kinase lack autophosphorylation activity.";
RL Arch. Biochem. Biophys. 383:233-237(2000).
DR EMBL; AF249317; AAF91336.1; -
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004040; STY_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00221; STYK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Kinase; Tyrosine-protein kinase.
SQ SEQUENCE 360 AA; 39375 MW; E170D5DD20E2D6F CRC64;
Query Match 77.4%; Score 41; DB 10; Length 360;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AOPGPVLS 8
Db 160 AOPGPVLS 167
|||||
RESULT 3
ID 09LKY3 PRELIMINARY; PRT; 360 AA.
AC 09LKY3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE PtlI kinase-like protein.
GN PtlIB.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE=21036115; PubMed=11185558;
RA Staswick P.;
RT "Two expressed soybean genes with high sequence identity to tomato
PtlI kinase lack autophosphorylation activity.";
RL Arch. Biochem. Biophys. 383:233-237(2000).
DR EMBL; AF249318; AAF91337.1; -
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004040; STY_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00221; STYK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Kinase; Tyrosine-protein kinase.
SQ SEQUENCE 360 AA; 39481 MW; 0B541A3D95C3B90 CRC64;
Query Match 77.4%; Score 41; DB 10; Length 360;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AOPGPVLS 8
Db 160 AOPGPVLS 167
|||||
RESULT 4
ID 09LUT0 PRELIMINARY; PRT; 364 AA.
AC 09LUT0;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE serine/threonine protein kinase-like protein (Hypothetical 39.6 kDa
protein).
GN At3G17410.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
features of the regions of 4,504,864 bp covered by sixty P1 and TAC
clones.";
RL DNA Res. 7:131-135(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Yamada K., Banh J., Chang C.H., Chang E., Dale J.M., Goldsmith A.D.,
Lee J.M., Ondera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
Yamamura Y., Yu G., Yu S., Bowser L., Carlnict P., Chen H., Cheuk R.,
Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
Kawai J., Kim C., Lam B., Lin J., Meyers M.C., Miranda M.,
Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
Theologis A.;
RT "Full length cDNA of gene At3g17410 (GI:15229035)."
RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AB022216; BAB02745.1; -
DR EMBL; AY070413; AAL49909.1; -
DR HSSP; P11362; IFC1
InterPro; IPR000719; Euk_pkinase.

DR InterPro: IPR002290; Ser_thr-kinase.
DR InterPro: IPR004040; STY_pkinase.
DR InterPro: IPR001245; TYR_pkinase.
DR Pfam: PF00069; pkinase.1.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_pkinase.1.
DR SMART: SM00221; STYKc.1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR ATP-binding: Kinase: Serine/threonine-protein kinase; Transferase;
KW Hypothetical protein
SQ SEQUENCE 364 AA; 39562 MW; 62A7FE8F249F332E CRC64;

Query Match 77.4%; Score 41; DB 10; Length 364;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AQPpVLSP 8
DB 162 AQPpVLSP 169

RESULT 5
080719
ID 080719 PRELIMINARY; PRT; 365 AA.
AC 080719;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative pto kinase (Hypothetical 39.9 kDa protein).
GN F14M.11 OR AT2G47060.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
RA Shen M., Ronning C.M., Fraser C.M., Somerville C.R., Venter J.C.,
RT "Arabidopsis thaliana chromosome II BAC F14M4 genomic sequence.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Theologis A., Ecker J., Davis R.W.,
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Arabidopsis full length cDNA clones.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL: AC004411; AAC34243.1; -;
DR EMBL: AF325090; AK17158.1; -;
DR EMBL: AY080876; AAL87347.1; -;
DR HSSP: P1362; 1FGI.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002290; Ser_thr-kinase.
DR InterPro: IPR004040; STY_pkinase.
DR Pfam: PF00069; pkinase.1.
DR ProDom: PD000001; Euk_pkinase.1.
DR SMART: SM00221; STYKc.1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.

DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
KW Hypothetical protein; ATP-binding; Serine/threonine-protein kinase;
KW Transferase.
SQ SEQUENCE 365 AA; 39931 MW; ACAFECC6C79E77CAB CRC64;

Query Match 77.4%; Score 41; DB 10; Length 365;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AQPpVLSP 8
DB 166 AQPpVLSP 173

RESULT 6
09U191
ID 09U191 PRELIMINARY; PRT; 258 AA.
AC 09U191;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Hypothetical transmembrane protein L302.06.
GN L302.06.
OS Leishmania major.
OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIDLIN;
RA Ivens A.C., Murphy L., Quail M., Harris D., Oliver K., Lawson D.,
RA Rajandream M.A., Barrell B.G.,
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIDLIN;
RA MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
DR EMBL: AL389894; CAC22674.1; -;
KW Transmembrane.
SQ SEQUENCE 258 AA; 27729 MW; 1F214B2109E0D2DB CRC64;

Query Match 73.6%; Score 39; DB 5; Length 258;
Best Local Similarity 70.0%; Pred. No. 19;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 AQPpVLSP 10
DB 211 AQPpVLSP 220

RESULT 7
09UGS3
ID 09UGS3 PRELIMINARY; PRT; 797 AA.
AC 09UGS3;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE D1756G23.1 (Novel leucine rich protein) (Fragment).
GN D1756G23.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Corby N.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL035681; CAB63072.1; -;
DR InterPro: IPR001611; LRR.
DR InterPro: IPR000483; LRR_Cterm.

DR InterPro: IPR000372; LRR_Nterm.
 DR InterPro: IPR003592; LRR_out.
 DR InterPro: IPR003591; LRR_Cyp.
 DR Pfam: PF00560; LRR_18.
 DR Pfam: PF01463; LRRCT_2.
 DR Pfam: PF01462; LRRNT_1.
 DR PRINTS: PR00019; LEURICHRPT.
 DR SMART: SM00370; LRR_8.
 DR SMART: SM00082; LRRCT_2.
 DR SMART: SM00013; LRRNT_2.
 DR SMART: SM00369; LRR_TYP_4.
 DR NON_TER 797
 FT SEQUENCE 797 AA; 85614 MW; 8C3247883EAE59AD CRC64;

Query Match 73.6%; Score 39; DB 4; Length 797;
 Best Local Similarity 88.9%; Pred. No. 59;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACPGVLSQ 9
 DB 261 ALPGVLSQ 269

RESULT 8

ID 08WV2 PRELIMINARY; PRT; 244 AA.
 AC 08WV2: 01-MAR-2002 (TREMblrel. 20, Created)
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Hypothetical 27.2 kDa protein.
 GN P14776
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Zhang P.P., Zhou X.M., Jiang H.Q., Huang Y., Qin W.X., Zhao X.T.,
 Wan D.F., Gu J.R.:
 RT "Novel human cDNA clones with function of inhibiting cancer cell
 growth."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF181347; AAL55854.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 244 AA; 27187 MW; 5E7520DB8245F8AB CRC64;

Query Match 71.7%; Score 38; DB 4; Length 244;
 Best Local Similarity 66.7%; Pred. No. 28;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 QPGPVLSQ 10
 DB 209 QPGPLKRP 217

RESULT 9

ID 093GV7 PRELIMINARY; PRT; 265 AA.
 AC 093GV7:
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Thioesterase.
 OS Streptomyces avermiltis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=33903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=21477403; PubMed=11572948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,

RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
 RT "Genome sequence of an industrial microorganism Streptomyces
 RT avermiltis: Deducing the ability of producing secondary
 RT metabolites."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
 DR EMBL: AB070956; BAB69404.1; -
 DR InterPro: IPR001031; Thioesterase.
 DR Pfam: PF00975; Thioesterase; 1.
 SQ SEQUENCE 265 AA; 28688 MW; 70DB99578E56C48A CRC64;

Query Match 71.7%; Score 38; DB 2; Length 265;
 Best Local Similarity 87.5%; Pred. No. 30;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 PGPVLSQ 10
 DB 194 PGPVLSQ 201

RESULT 10

ID 096JW9 PRELIMINARY; PRT; 397 AA.
 AC 096JW9:
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE CDNA FLJ14925 f1s, clone PLACE1008643, weakly similar to
 DE Inter-alpha-trypsin inhibitor heavy chain H2 precursor.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PLACENTA;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Wagatsuna M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murekami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,
 RA Niomiya K., Iwayanagi T.;
 RT "NEO human cDNA sequencing project."
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK027831; BAB55397.1; -
 DR InterPro: IPR002035; VWF_A.
 DR Pfam: PF00092; vwa; 1.
 DR PROSITE: PS50234; VWF_A; 1.
 SQ SEQUENCE 397 AA; 44345 MW; 096673471189AF75 CRC64;

Query Match 71.7%; Score 38; DB 4; Length 397;
 Best Local Similarity 66.7%; Pred. No. 45;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 QPGPVLSQ 10
 DB 353 QPGPLKRP 361

RESULT 11

ID 096JV8 PRELIMINARY; PRT; 397 AA.
 AC 096JV8:
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE CDNA FLJ14943 f1s, clone PLACE1011371, weakly similar to
 DE Inter-alpha-trypsin inhibitor heavy chain H2 precursor.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]

```

RP SEQUENCE FROM N.A.
RC TISSUE-PLACENTA;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y., Oshima A.;
RT "NDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK027849; BAB55409.1; -
DR InterPro: IPR002035; WVF_A.
DR Pfam: PF00092; WVF.1.
DR PROSITE: PSS0234; WVF.A.1.
SQ SEQUENCE 397 AA; 44342 MW; 9841AA56E65AF64 CRC64;

Query Match 71.7%; Score 38; DB 4; Length 397;
Best Local Similarity 66.7%; Pred. No. 45;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QPGPVLSOP 10
Db 353 QPGPLTKRP 361

RESULT 12
Q8XUK6 PRELIMINARY; PRT; 583 AA.
ID 08XUK6
AC Q8XUK6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Probable activation/secretion signal peptide protein.
GN RSC3182 OR RS06080.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GM11000;
RX MEDLINE-21681879; PubMed-11823852;
RA Salanoubat M., Genin S., Artiguenave F., Guzy J., Mangenot S.,
RA Ariat M., Billault A., Brotlier P., Camus J.C., Catolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Sigulier P., Thebault P., Whalen M., Winkler P., Levy M.,
RA Weissbach J., Boucher C.A.;
RL "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL: AL646074; CAD16970.1; -
RW Complete proteome.
SQ SEQUENCE 583 AA; 62313 MW; CEDE73F9943A5573 CRC64;

Query Match 71.7%; Score 38; DB 16; Length 583;
Best Local Similarity 70.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AOPGPVLSOP 10
Db 26 AOPGPVLEDP 35

RESULT 13
Q8XPV1 PRELIMINARY; PRT; 583 AA.
ID 08XPV1
AC Q8XPV1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Probable activation/secretion signal peptide protein.
GN RSP1535 OR RS06081.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
SQ Plasmid megaplasmid.

```

```

OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GM11000;
RX MEDLINE-21681879; PubMed-11823852;
RA Salanoubat M., Genin S., Artiguenave F., Guzy J., Mangenot S.,
RA Ariat M., Billault A., Brotlier P., Camus J.C., Catolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Sigulier P., Thebault P., Whalen M., Winkler P., Levy M.,
RA Weissbach J., Boucher C.A.;
RL "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL: AL646085; CAD18686.1; -
RW Complete proteome.
SQ SEQUENCE 583 AA; 62554 MW; C248BDA5001DF3C3 CRC64;

Query Match 71.7%; Score 38; DB 16; Length 583;
Best Local Similarity 70.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AOPGPVLSOP 10
Db 26 AOPGPVLEDP 35

RESULT 14
Q99LJ9 PRELIMINARY; PRT; 649 AA.
ID Q99LJ9
AC Q99LJ9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical 69.6 kDa protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC003195; AAH03195.1; -
DR InterPro: IPR000345; Cytochrome_b.
DR InterPro: IPR000571; Znf_CCH.
DR InterPro: IPR001841; Znf_CCH.
DR Pfam: PF00642; Zf-CCH; 2.
DR SMART: SM00184; RING; 1.
DR SMART: SM00356; Znf_C3H1; 3.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
RW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 649 AA; 69599 MW; C6B161FB791D2CEB CRC64;

Query Match 71.7%; Score 38; DB 11; Length 649;
Best Local Similarity 77.8%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QPGPVLSOP 10
Db 178 QPGPVLYMP 186

RESULT 15
Q9BT94 PRELIMINARY; PRT; 728 AA.
ID Q9BT94
AC Q9BT94;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical 80.6 kDa protein.

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKIN;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004282; AAH04282.1; -
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00092; vwa; 1.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWF_A; 1.
KW Hypothetical protein.
SQ SEQUENCE 728 AA; 80604 MW; 118A1B32F89BCECA CRC64;

Query Match 71.7%; Score 38; DB 4; Length 728;
Best Local Similarity 66.7%; Pred. No. 83;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 2 QPGPVLSOP 10
||||:|:|
Db 444 QPGPILKRP 452

Search completed: February 16, 2003, 21:59:13
Job time : 3.70236 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:40:36 ; Search time 8.43255 Seconds
(without alignments)
1738.213 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_552
Perfect score: 562
Sequence: 1 ACPGVLSPAGIPGTSSSK.....ASPTQNRKIKENTSCFRDE 110

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	562	100.0	801	19 AAW31494	Human hRAFI105 pr
2	562	100.0	801	22 ABB03190	Rat E1f-1 polymorp
3	562	100.0	852	21 AAB50171	Human transcriptio
4	92.5	16.5	737	15 AAB56494	Human polyepitide
5	92.5	16.5	737	17 AAB56494	Human polyepitide
6	92.5	16.5	737	18 AAB56494	Human polyepitide
7	79.5	14.1	3726	22 ABB63947	Human polyepitide
8	78.5	14.0	622	22 ABB63947	Human polyepitide
9	76	13.5	666	22 AAB94560	Human protein sequ
10	74	13.2	524	21 AAB03192	Rat E1f-1 polymorp

11	74	13.2	579	21 AAB03191	Rat E1f-1 polymorp
12	74	13.2	615	21 AAB03190	Rat E1f-1 polymorp
13	73	13.0	591	22 AAB50171	Human transcriptio
14	73	13.0	595	22 AAB38685	Human polyepitide
15	73	13.0	619	22 AAB38685	Human polyepitide
16	73	13.0	625	22 AAB40471	Human polyepitide
17	73	13.0	625	22 AAB40472	Human polyepitide
18	72.5	12.9	412	23 ABB38928	Human polyepitide
19	72.5	12.9	1236	22 ABB08165	Staphylococcus epl
20	72	12.8	2296	22 ABB59050	Novel human diagno
21	72	12.8	2639	22 ABB15016	Drosophila melanog
22	71.3	12.7	979	11 AAB08338	Novel human diagno
23	71	12.6	528	11 AAB98977	Complete form of h
24	70.5	12.5	128	22 AAB46608	Human PCPL. Homo
25	69.5	12.4	1045	22 ABB08332	Novel human diagno
26	69.5	12.4	1532	21 AAB40945	Human OREF ORF709
27	69	12.3	881	22 ABB05595	Novel human diagno
28	69	12.3	1420	22 ABB63410	Drosophila melanog
29	68.5	12.2	215	22 AAB91954	C glutamylcum prote
30	68	12.1	1023	23 AAB82954	Human homologue of
31	67.5	12.0	736	23 ABB35614	Fungal ZBC protein
32	67.5	12.0	1215	23 ABB97489	Novel human protei
33	67.5	12.0	2442	22 ABB44556	Human wound healin
34	67	11.9	363	22 ABB61938	Drosophila melanog
35	67	11.9	375	22 AAB03535	Enterococcus faeca
36	67	11.9	725	22 AAB95238	Human protein sequ
37	67	11.9	748	22 AAB37199	Staphylococcus aur
38	67	11.9	792	22 AAB36734	Staphylococcus aur
39	67	11.9	835	23 ABB05681	Human differentiat
40	67	11.9	1532	22 ABB68585	Drosophila melanog
41	67	11.9	1694	21 AAB48461	Rat laminin 5 poly
42	67	11.9	1725	21 AAB48460	Rat laminin 5 poly
43	67	11.9	3169	22 ABB64569	Drosophila melanog
44	66.5	11.8	355	22 AAB05117	Human deaminase en
45	66.5	11.8	355	23 AAB05118	Human mutant deam

ALIGNMENTS

RESULT 1
ID AAW31494 standard; Protein: 801 AA.
AAW31494:
AC AAW31494:
XX
XX
DT 28-APR-1998 (first entry)
XX
XX Human hRAFI105 protein.
DE
XX
XX TATA-binding protein associated binding factor 105; human; activator;
KW hRAFI105; transcription factor; TRFID; transcriptional activation;
KW antibodies; diagnosis; therapy; biopharmaceutical industry.
XX
XX Homo sapiens.
OS
XX
XX US5710025-A.
PN
XX
XX 20-JAN-1998.
PD
XX
XX 02-OCT-1996; 96US-0725012.
PF
XX
XX 02-OCT-1996; 96US-0725012.
PR
XX
XX (REGC) UNIV CALIFORNIA.
PA
XX
XX Dikstein R, Tjian R;
XX
XX WPI: 1998-109818/10.
DR
XX
XX N-PSDB: AAV02872.
XX
XX DNA encoding human tata-binding protein associated factor - for
PT
XX
XX producing recombinant protein

XX Claim 1; Col 17-22; 12pp; English.

CC This cDNA sequence represents a human cell-binding protein associated
 CC factor, hTAFII105, isolated from Daudi cell nuclear extracts. Tightly
 CC associated subunits (TAF's) are components of the transcription factor
 CC TFIID and are thought to mediate transcriptional activation. This encoded
 CC protein may be produced recombinantly from transformed host cells or
 CC purified from human cells. hTAFII105 specific binding agents such as
 CC specific antibodies could be used for diagnosis (e.g. genetic
 CC hybridisation screens for hTAFII105 transcripts), therapy (e.g. gene
 CC therapy to modulate hTAFII105 gene expression) and in the
 CC biopharmaceutical industry (e.g. as immunogens, reagents for isolating B
 CC cell specific activators or other transcriptional regulators).

XX Sequence 801 AA:

SQ

Query Match 100.0%; Score 562; DB 19; Length 801;
 Best Local Similarity 100.0%; Pred. No. 1.6e-54;
 Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AOPGPVLSQPAIGPTGSSSKQLFSLFHVYQOPSGNEKQVTTISHSTLTIOKCGCKTMP 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 443 AOPGPVLSQPAIGPTGSSSKQLFSLFHVYQOPSGNEKQVTTISHSTLTIOKCGCKTMP 502

OY 61 VNTIIPTSQFPASIIKQITLPGNKILSLQASPTQGNRIKENVTSQFRDE 110
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 503 VNTIIPTSQFPASIIKQITLPGNKILSLQASPTQGNRIKENVTSQFRDE 552

RESULT 2
 ABG09468

ID ABG09468 standard; Protein; 801 AA.

XX ABG09468;

AC

XX 13-FEB-2002 (first entry)

DT

XX

DE Novel human diagnostic protein #9459.

XX

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

DR N-PSDB; AAS73655.

XX

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

XX

PS Claim 20; SEQ ID No 39827; 103pp; English.

XX

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 801 AA:

SQ

Query Match 100.0%; Score 562; DB 22; Length 801;
 Best Local Similarity 100.0%; Pred. No. 1.6e-54;
 Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AOPGPVLSQPAIGPTGSSSKQLFSLFHVYQOPSGNEKQVTTISHSTLTIOKCGCKTMP 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 443 AOPGPVLSQPAIGPTGSSSKQLFSLFHVYQOPSGNEKQVTTISHSTLTIOKCGCKTMP 502

OY 61 VNTIIPTSQFPASIIKQITLPGNKILSLQASPTQGNRIKENVTSQFRDE 110
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 503 VNTIIPTSQFPASIIKQITLPGNKILSLQASPTQGNRIKENVTSQFRDE 552

RESULT 3
 AAY57279

ID AAY57279 standard; Protein; 852 AA.

XX AAY57279;

AC

XX 06-JUN-2000 (first entry)

DT

XX

DE Transcription factor subunit TAFII105 polypeptide.

XX

XX TATA box-binding protein associated factor II 105; TAFII105; cancer;
 KW transcription factor; apoptosis; cytostatic; immunosuppressive;
 KW antiinflammatory; virucide; antibacterial.

XX

OS Homo sapiens.

XX

OS

PH Key Location/Qualifiers

FT Misc-difference 802

FT /label= GLX

FT /note= "encoded by TGA"

FT Misc-difference 834

FT /label= GLX

FT /note= "encoded by TAA"

XX

PN WO200012699-A1.

XX

PD 09-MAR-2000.

XX

PF 25-AUG-1999; 99WO-IL00464.

XX

PR 27-AUG-1998; 98IL-0125971.

XX

PA (YEDA) YEDA RES & DEV CO LTD.

XX

PI Dikstein R, Yamit-hezi A;

XX

XX WPI; 2000-256640/22.

DR N-PSDB; AA290465.

XX

XX Polypeptide encoding TATA box binding protein associated factor II 105
 PT useful for treating e.g. cancers and inducing apoptosis has a dominant

PT negative effect on the normal biological activity of the binding protein -

PS Claim 7: Fig 2: 48bp: English.

CC This represents a polypeptide comprising a (modified) fragment (I) of a TATA box-binding protein associated factor II 105 (TAFII105). A pharmaceutical composition comprising (I) or the polynucleotide or an inhibitor or antagonist of (I) is useful for treating cancers and inducing apoptosis in pathological cells. The composition is also useful for treating autoimmune diseases, inflammatory processes and viral or bacterial infections.

CC Sequence 852 AA;

Query Match 100.0%; Score 562; DB 21; Length 852;
Best Local Similarity 100.0%; Pred. No. 1.8e-54;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AOPGVLSQAPAGIPITGSSSKQLFSLFHVVOQPSGNEKOVTTISHSSTLTIOKCGQKTM 60
DB 443 AOPGVLSQAPAGIPITGSSSKQLFSLFHVVOQPSGNEKOVTTISHSSTLTIOKCGQKTM 502
QY 61 VNTIIPTSOPPPASIIKQITLPGKIIISLQASPTQKNRIKENTSCFRDE 110
DB 503 VNTIIPTSOPPPASIIKQITLPGKIIISLQASPTQKNRIKENTSCFRDE 552

RESULT 4
ID AAR56494 standard; Protein: 737 AA.
AC AAR56494;
DT 23-MAR-1995 (first entry)
XX
DE TATA-binding protein-associated factor hTAFII130.
KW TATA-binding protein associated factor; hTAFII130; screening;
KM diagnostic; therapeutic; gene transcription regulation.
XX
OS Homo sapiens.
XX
PN WO9417087-A.
PD 04-AUG-1994.
XX
PF 28-JAN-1994; 94NO-US01114.
XX
PR 28-JAN-1993; 93US-0013412.
PR 30-JUN-1993; 93US-0087119.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Comal L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
PI Tjian R, Wang E, Weinzierl ROJ;
XX
DR WPI: 1994-264019/32.
DR N-PSDB; AAO70731.
XX
PT TATA-binding protein associated protein factors - and
PT corresponding nucleotide sequence and deriv. antibodies, useful
PT in screening, diagnostics and therapeutics
XX
PS Disclosure: Page 142; 180pp; English.
XX
CC The TATA-binding protein associated factor hTAFII130 (including
CC specific antibodies and fusion products) are used in drug screening,
CC diagnostics and therapeutics. They are used in the development of
CC specific biochemical assays for screening compounds that agonise or
CC antagonise selected transcription factors involved in regulating
CC gene expression associated with human pathology.
XX

SQ Sequence 737 AA;
Query Match 16.5%; Score 92.5; DB 15; Length 737;
Best Local Similarity 27.3%; Pred. No. 0.18;
Matches 33; Conservative 19; Mismatches 50; Indels 19; Gaps 5;

QY 2 OPGVLS--OPAGIPITGSSSKQLFSLFHVVOQPSGNEKOVTTISHSSTLTIOKCGQKTM 59
DB 370 QP-VLSLTQPTQGVGKQGPPLVLIQOPPKPGALIRPVQVTLQPMVALRQ----- 422
QY 60 PVNIIIP-----SQPPASIIKQITLPGKIIISLQASPTQKNRIKENTSCFRD 109
DB 423 PNNRIMLTTPQQLQNLNPLQPVVPAVLPETKALSAVSAQAAAAQKNKLEKGGSGFRD 482
QY 110 E 110
DB 483 D 483

RESULT 5
ID AAM06084 standard; Protein: 737 AA.
AC AAM06084;
DT 27-JAN-1997 (first entry)
XX
DE Human TATA-binding protein associated factor hTAFII130 protein.
KW Human; TATA-binding protein; TBP associated factor; TFIID; holoenzyme;
KW RNA polymerase II; transcription; messenger RNA; nuclear fraction;
KW lambda-gli1; expression library.
XX
OS Homo sapiens.
XX
PN US5534410-A.
PD 09-JUL-1996.
XX
PF 28-JAN-1993; 93US-0013412.
XX
PR 28-JAN-1994; 94US-0188582.
PR 28-JAN-1993; 93US-0013412.
PR 30-JUN-1993; 93US-0087119.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Comal L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
PI Tjian R, Wang E, Weinzierl ROJ;
XX
DR WPI: 1996-333245/33.
DR N-PSDB; AAT42217.
XX
PT Screen for cpds. that bind human TATA-binding protein associated
PT factor - by testing ability to bind to polypeptide fragments of the
PT factor, useful as (ant)agonists of transcription factors involved in
PT disease.
XX
PS Examples: Column 105-112; 86pp; English.
XX
CC This is the amino acid sequence of the human TATA-binding protein (TBP)
CC associated factor (TAF) designated TAFII130. The protein is a component
CC of the TFIID fraction required for reconstituting RNA polymerase II in
CC vitro transcription activity. The encoded protein has an estimated mol.
CC wt. of 130 kD by SDS-PAGE.
CC The invention relates to purified proteins involved in transcription
CC by RNA polymerase II, the RNA polymerase which transcribes messenger
CC RNA. RNA polymerase II transcription proceeds in vitro upon addition
CC of several nuclear fractions designated TFIID, B, D, E, F, H, I and J
CC to RNA polymerase II holoenzyme. Fraction TFIID has been shown to
CC contain a TBP and other TAFs. Purification of TFIID and separation of
CC its components reveals 7 proteins ranging in size from 30-250 kD.
CC Serum raised against the TFIID fraction allowed cloning of the corresp.

CC genes from lambda-gt11 expression libraries.
 XX
 SQ Sequence 737 AA;
 Query Match 16.5%; Score 92.5; DB 17; Length 737;
 Best Local Similarity 27.3%; Pred. No. 0.18;
 Matches 33; Conservative 19; Mismatches 50; Indels 19; Gaps 5;
 QY 2 QPEPVLS--QPAQIPMGSSSKQLFSLFHYVQOPSGGNGKQVTTISSTLTTCCKQKTM 59
 DB 370 QP-PVLSLQPTQVGVGKQGOPTPLVIOQPPKGCALIRPPQVTLTQTVPVALRQ----- 422
 QY 60 PVTITPT-----SQFPASILKQITLPGNKIL--SIQASPTQKNRIKENVTSQCFRD 109
 DB 423 PNRRLMTTPQQTQLNPLQPPVYKPAVLPGTKALSAVSAQAQAQKNKILKEPGGGSFRD 482
 QY 110 E 110
 DB 483 D 483
 RESULT 6
 ID AAM25019 standard; Protein: 737 AA.
 XX AAM25019;
 AC AAM25019;
 DT 08-OCT-1997 (first entry)
 DE TATA-binding protein associated factor, hTAFII130.
 XX
 DE TATA-binding protein associated factor; TAF; nuclear protein;
 KM RNA polymerase transcription; TATA-binding protein; TBP;
 KW Initiation.
 XX
 OS Homo sapiens.
 XX
 PN US5637686-A.
 PD 10-JUN-1997.
 XX
 PF 28-JAN-1993; 93US-0013412.
 XX
 PR 28-JAN-1994; 94US-0188582.
 PR 28-JAN-1993; 93US-0013412.
 PR 30-JUN-1993; 93US-0087119.
 PR 09-MAY-1996; 96US-0646715.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Comat L, Dynalact BD, Hoey T, Ruppert S, Tanese N;
 PI Tjian R, Wang E, Weinzierl ROJ;
 DR WPI: 1997-319113/29.
 DR N-PSDB: AAT79595.
 XX
 PT Nucleic acids encoding human TATA-binding protein associated factor
 PT (TAF) peptide(s) - for production of recombinant peptide(s), used
 PT for modulating transcription of TAFs
 XX
 PS Claim 1; Column 111-116; 86pp; English.
 CC AAM25018 represents TATA-binding protein associated factor (TAF)
 CC polypeptide, hTAFII100 (mol. weight 100kD). TAF peptides derived
 CC from hTAFII30 alpha, hTAFII30 beta, hTAFII40, hTAFII70, hTAFII100,
 CC hTAFII130, hTAFII250, hTAFII48 and hTAFII10 and nucleic acids encoding
 CC them, are used to modulate transcription, including transcription
 CC Initiation. TAFs are nuclear proteins involved in RNA polymerase I,
 CC II and III transcription. The peptides act by binding to a different
 CC TAF, an activator, or TBP (TATA-binding protein) or competitively
 CC inhibiting association of a TAF domain with another compound, typically
 CC a protein like TBP or another TAF, an activator, or DNA.

SQ Sequence 737 AA;
 Query Match 16.5%; Score 92.5; DB 18; Length 737;
 Best Local Similarity 27.3%; Pred. No. 0.18;
 Matches 33; Conservative 19; Mismatches 50; Indels 19; Gaps 5;
 QY 2 QPEPVLS--QPAQIPMGSSSKQLFSLFHYVQOPSGGNGKQVTTISSTLTTCCKQKTM 59
 DB 370 QP-PVLSLQPTQVGVGKQGOPTPLVIOQPPKGCALIRPPQVTLTQTVPVALRQ----- 422
 QY 60 PVTITPT-----SQFPASILKQITLPGNKIL--SIQASPTQKNRIKENVTSQCFRD 109
 DB 423 PNRRLMTTPQQTQLNPLQPPVYKPAVLPGTKALSAVSAQAQAQKNKILKEPGGGSFRD 482
 QY 110 E 110
 DB 483 D 483
 RESULT 7
 ID ABB63947 standard; Protein: 3726 AA.
 XX ABB63947;
 AC ABB63947;
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 18633.
 XX
 DE Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 KW
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EW;
 DR WPI: 2001-656860/75.
 DR N-PSDB: ABL08050.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure: SEQ ID NO 18633; 21pp + Sequence Listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (AB57737-AB872072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 XX
 SQ Sequence 3726 AA;
 Query Match 14.1%; Score 79.5; DB 22; Length 3726;
 Best Local Similarity 18.2%; Pred. No. 51;
 Matches 30; Conservative 24; Mismatches 50; Indels 61; Gaps 5;


```

QY      3  PGPALSDPAGIPGSSSSSKOL--PSLPHVVOQPPSGNEKVO----- 40
      2953  PKVTISQOR-IPAOYQOQOQQOQMHHIPQOQOPLQOQOVVOQPSMPITTLAEAPVVSQ 30111
QY      41  -----TTISHSSTLTIOKCGQKMPVNTIIPTSQFPASITLQKITLPGNKI 86
      3012  FYMERQALBQOELANRQVQHSSTSSSSSSSCSLPYNVNVPMQOQAPSTSSSTTPTNRKV 30711
Db      87  LSLD-----ASPTQKRRIE-----NNTSCF 107
      3072  LPMQORQBPAPLSNECPVSSPTPPKRPVQEDPIIHQMTASVSKCY 3116

```

XX	RESULT 8
AC	ABB60384
XX	ID ABB60384 standard; Protein; 622 AA.
XX	ABB60384;
DT	26-MAR-2002 (first entry)
XX	Drosophila melanogaster polypeptide SEQ ID NO 7944.
DE	Drosophila; developmental biology; cell signalling; insecticide;
XX	pharmaceutical.
KW	Drosophila melanogaster.
XX	WO200171042-A2.
XX	27-SEP-2001.
XX	23-MAR-2001; 2001WO-US09231.
XX	23-MAR-2000; 2000US-191637P.
XX	11-JUL-2000; 2000US-0614150.
XX	(PEKE) PE CORP NY.
XX	Venter JC, Adams M, Li PMD, Myers EM;
XX	WPI: 2001-656860/75.
DR	N-PSDB; ABL04487.
XX	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signalling and cell-cell
PT	interactions -
XX	
XX	Disclosure; SEQ ID NO 7944; 21pp + Sequence Listing; English.
XX	
XX	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC	sequences (ABB57737-ABB72072),
CC	(ABB57737-ABB72072), and the encoded proteins
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
XX	Sequence 622 AA;
XX	

Query Match 14.0%; **Score** 78.5; **DB** 22; **length** 622;
Best Local Similarity 27.5%; **Pred.** No.5.6;
Matches 28; **Conservative** 19; **Mismatches** 40; **Indels** 15; **Gaps** 5;

```
OY      5 PYLSPACIPIGSSSKQ--LESLEHYVQQPSGGNEKOVTTISHSTLTITQCGCKOTMPYN    62
          |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
DB     509 PRLINVLGGSCGNKKMRALFENFSASSIDHEN-----VTITYPSSLA---ASNQTAAAN    560
```

```

Oy      63  TITP-TSQPPASILKQITLPGKNIISLQ-----ASLPQKNRI 99
          | | :|:|:|:| | :|:| |
Db      561  AIAPTVTITPSSVGKVTVPNPREFILKPAKFFVGPASANEV 602

```

RESULT 9
AAB94560
ID AAB94560 standard; Protein; 696 AA

AC	AAB94560;	.
XX		
DT	26-JUN-2001	(first entry)
XX		

DE	Human protein sequence SEQ ID NO:153332.
XX	
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy
YX	

OS	Homo sapiens.
XX	
PN	EP1074617-A2.
XY	

PB	07-FEB-2001.
XX	
PF	28-JUL-2000; 2000EP-0116126.
XX	

PR 29-JUL-1999; 99JP-0248036
PR 27-AUG-1999; 99JP-0300253
PR 11-JAN-2000; 2000JP-0118776
PR 02-MAY-2000; 2000JP-0183767
PR 09-JUN-2000; 2000JP-0241899
XX

PA (HELI-) HELIX RES INST

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5002
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -

PS Claim 8; SEQ ID 15332; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialized methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

\$Q\$	Sequence	696 AA;
	Query Match	13.5%; Score 76; DB 22; Length 696;
	Best Local Similarity	23.0%; Pred. No. 13;
Matches	37; Conservative	23; Mismatches 41; Indels 60; Gaps 8

```

QY 3 PGVYL-----SOPAGIPFGSSSKOLFSLFHVVOQPSGNN-----EKQVTT 42
DB 5 PGVGMGIRLPAPSKPSETPPSTSSSAFVWNPVIOAVSSSAVNTTQAPSLLSGASF 64
QY 43 ISHSSTLTID-----KCGQAKT-----MPVNT--IIPTS-----QFP--- 71
DB 65 VSOAGTTLTALISPEPQSFASKTGSEIKITYSSGCGPVGTASLIRPLOSASFALLQLPQK 124
QY 72 --PASILKQITLPGNKILSLQASPTQKNRIKENVTSCFDE 110
DB 125 PVPSSILQHYA-----SLQMKRESQNPQDKDETNISIKRE 158

RESULT 10
ID AAB03192 standard; Protein; 524 AA.
AC AAB03192;
DT 23-OCT-2000 (first entry)
DE Rat Elf-1 polymorphic variant #3.
KW Elf-1; transcriptional regulatory factor; rat; polymorphic variant;
KW immunological disease; allergic disease.
OS Rattus sp.
FH Key Location/Qualifiers
FT Misc-difference 144..160
FT /note="This region is apparently encoded by GGC"
XX
XX JP2000135088-A.
XX
XX 16-MAY-2000.
XX
XX 30-OCT-1998; 98JP-0309595.
XX
XX 30-OCT-1998; 98JP-0309595.
XX
XX (ASAK ) ASAH I BREWERIES LTD.
XX
XX WPI; 2000-425802/37.
XX
XX N-PSDB; AAA61647.
XX
XX A transcription regulating factor Elf-1 cDNA and its polymorphism -
XX
XX Claim 3; Page 17-18; 21pp; Japanese.
XX
XX The invention relates to three polymorphic variants of a rat
XX transcriptional regulatory factor Elf-1 (AAB03190-B03192) and nucleic
XX acids encoding them (AAA61645-A61647). The invention also encompasses
XX the use of Elf-1 for regulating the transcription of a gene in an
XX expression vector. The cDNA encoding rat Elf-1 variant #1 (AAA61645)
XX was isolated using PCR primers (AAA61648-A61649) based on mouse Elf-1
XX cDNA, and the rat Elf-1 variant #2 and #3 cDNAs (AAA61646-A61647) were
XX subsequently isolated. The Elf-1 cDNA and protein sequences are
XX potentially useful for the prevention and treatment of immunological and
XX allergic diseases mediated by Elf-1. The present sequence represents
XX a variant (#3) of rat Elf-1.
XX
XX Sequence 524 AA.
XX
XX Query Match 13.2%; Score 74; DB 21; Length 524;
XX Best Local Similarity 30.9%; Pred. No. 14;
XX Matches 30; Conservative 11; Mismatches 38; Indels 18; Gaps 5;

QY 2 QPGPVLS--QPGAGIPFGSSSKOLFSLFHVVOQPSGNGEKQVTTISSTLTIDKCGQKTM 59
DB 271 QPSEVLRVQPSQAPYPT--QLFRVHVVOQVQAIPPEATIT--STMQEEMAGSSVQ 323
QY 60 PVNTIIPTSQFPASILKQITLPGNKIL--SLQASP 93

```

```

DB 324 GIRTQASTQVP-----VWSPGNGOQLHTVTTLQTPV 354

RESULT 11
ID AAB03191 standard; Protein; 579 AA.
AC AAB03191;
DT 23-OCT-2000 (first entry)
DE Rat Elf-1 polymorphic variant #2.
KW Elf-1; transcriptional regulatory factor; rat; polymorphic variant;
KW immunological disease; allergic disease.
OS Rattus sp.
FH Key Location/Qualifiers
FT Misc-difference 133
FT /note="Encoded by GAG"
FT Misc-difference 149
FT /note="Encoded by GAR"
XX
XX JP2000135088-A.
XX
XX 16-MAY-2000.
XX
XX 30-OCT-1998; 98JP-0309595.
XX
XX 30-OCT-1998; 98JP-0309595.
XX
XX (ASAK ) ASAH I BREWERIES LTD.
XX
XX WPI; 2000-425802/37.
XX
XX N-PSDB; AAA61646.
XX
XX A transcription regulating factor Elf-1 cDNA and its polymorphism -
XX
XX Claim 2; Page 13-14; 21pp; Japanese.
XX
XX The invention relates to three polymorphic variants of a rat
XX transcriptional regulatory factor Elf-1 (AAB03190-B03192) and nucleic
XX acids encoding them (AAA61645-A61647). The invention also encompasses
XX the use of Elf-1 for regulating the transcription of a gene in an
XX expression vector. The cDNA encoding rat Elf-1 variant #1 (AAA61645)
XX was isolated using PCR primers (AAA61648-A61649) based on mouse Elf-1
XX cDNA, and the rat Elf-1 variant #2 and #3 cDNAs (AAA61646-A61647) were
XX subsequently isolated. The Elf-1 cDNA and protein sequences are
XX potentially useful for the prevention and treatment of immunological and
XX allergic diseases mediated by Elf-1. The present sequence represents
XX a variant (#2) of rat Elf-1.
XX
XX Sequence 579 AA.
XX
XX Query Match 13.2%; Score 74; DB 21; Length 579;
XX Best Local Similarity 30.9%; Pred. No. 16;
XX Matches 30; Conservative 11; Mismatches 38; Indels 18; Gaps 5;

QY 2 QPGPVLS--QPGAGIPFGSSSKOLFSLFHVVOQPSGNGEKQVTTISSTLTIDKCGQKTM 59
DB 326 QPSEVLRVQPSQAPYPT--QLFRVHVVOQVQAIPPEATIT--STMQEEMAGSSVQ 378
QY 60 PVNTIIPTSQFPASILKQITLPGNKIL--SLQASP 93
DB 379 GIRTQASTQVP-----VWSPGNGOQLHTVTTLQTPV 409

RESULT 12
ID AAB03190 standard; Protein; 615 AA.
XX

```

AC AAB03190;
XX 23-OCT-2000 (first entry)
XX
XX Rat E1f-1 polymorphic variant #1.
DE
XX E1f-1; transcriptional regulatory factor; rat; polymorphic variant;
KW immunological disease; allergic disease.
XX
XX Rattus sp.
OS
XX JP2000135088-A.
PN
XX 16-MAY-2000.
PD
XX 30-OCT-1998; 98JP-0309595.
PE
XX 30-OCT-1998; 98JP-0309595.
PR
XX (ASAK) ASAMI BREWERIES LTD.
PA
XX WPI: 2000-425802/37.
DR N-PSDB: AAB61645.
XX
XX A transcription regulating factor E1f-1 cDNA and its polymorphism -
PT
XX Claim 1; Page 8-10; 21pp; Japanese.
PS
XX The invention relates to three polymorphic variants of a rat
CC transcriptional regulatory factor E1f-1 (AAB03190-B03192) and nucleic
CC acids encoding them (AAB61645-A61647). The invention also encompasses
CC the use of E1f-1 for regulating the transcription of a gene in an
CC expression vector. The cDNA encoding rat E1f-1 variant #1 (AAB61645)
CC was isolated using PCR primers (AAB61648-A61649) based on mouse E1f-1
CC cDNA, and the rat E1f-1 variant #2 and #3 cDNAs (AAB61646-A61647) were
CC subsequently isolated. The E1f-1 cDNA and protein sequences are
CC potentially useful for the prevention and treatment of immunological and
CC allergic diseases mediated by E1f-1. The present sequence represents
CC a variant (#1) of rat E1f-1.
XX
SQ Sequence 615 AA;
Query Match 13.2%; Score 74; DB 21; Length 615;
Best Local Similarity 30.9%; Pred. No. 18;
Matches 30; Conservative 11; Mismatches 38; Indels 18; Gaps 5;
QY 2 QPGVPLS--QPGAGPTGSSSKQLSLFHVVO--PSGNGKQVTTISHSTLTIOKCGKTM 59
DB 362 QPSEVLRTVQPSQAPVPT--QLFRVHVVPVQVAIPEARIT---STMOEFAAGSSVQ 414
QY 60 PVNTIIPTSOPFPASILKQITLPGNKIL---SLQASP 93
DB 415 GIRTQASTQV-----VVVSPGNQLHTVLTQIVP 445
RESULT 13
AAB50171
ID AAB50171 standard; Protein: 591 AA.
XX
AC AAB50171;
XX
XX 05-FEB-2002 (first entry)
XX
XX Human transcription factor TRFX-22.
DE
XX Human; transcription factor; TRFX; cell proliferative disease;
KW autoimmune disease; inflammation; neurological disease;
KW developmental disorder; cancer; AIDS; infection; cytostatic; anti-HIV;
KW neuroprotective; anti-inflammatory; gene therapy.
XX
XX Homo sapiens.
XX
XX W0200172777-A2.
PN

XX 04-OCT-2001.
PD
XX 13-MAR-2001; 2001WO-US08117.
PE
XX 13-MAR-2000; 2000US-0188986.
PR
XX (INCY-) INCYTE GENOMICS INC.
PA
XX Hillman JI, Baughn MR, Yue H, Lal P, Lu DAM, Patterson C;
PI Azimzai Y, Bandman O, Tang YT, Mathur P, Shah P, Au-Young J;
PI Reddy R;
XX WPI: 2001-570896/64.
DR N-PSDB: ABA82995.
XX
XX Novel transcription factor polypeptides, used to treat diseases
PT associated with altered activity and expression of TRFX, and to screen
PT for agents capable of modulating its activity -
PS
XX Claim 1; Pages 164-166; 327pp; English.
XX
XX The present sequence is the protein sequence for a human transcription
CC factor. The transcription factor and its coding sequence are useful in
CC the diagnosis, treatment and prevention of diseases associated with
CC altered expression of the transcription factor e.g. cell proliferative,
CC autoimmune/inflammatory, neurological and developmental disorders. A
CC number of specific disorders/diseases are given in the specification,
CC including: arteriosclerosis, cirrhosis, hepatitis, cancers, AIDS,
CC allergies, anaemia, asthma, autoimmune thyroiditis, bronchitis, atopic
CC dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,
CC Grave's disease, multiple sclerosis, osteoarthritis, pancreatitis,
CC psoriasis, rheumatoid arthritis, systemic lupus erythematosus, ulcerative
CC colitis, uveitis, Alzheimer's disease, Huntington's disease, Parkinson's
CC disease, stroke, and viral, bacterial, fungal and protozoal infections.
XX
SQ Sequence 591 AA;
Query Match 13.0%; Score 73; DB 22; Length 591;
Best Local Similarity 28.9%; Pred. No. 22;
Matches 37; Conservative 11; Mismatches 50; Indels 30; Gaps 6;
QY 1 AOPGPVLS--OPAGIPTGSSSKQLSLFHVVO--PSGNGKQVTTISHSTLTIOK 54
DB 324 AOPSEVLRTVQPSQAPVPT--QLFRVHVVPVQVAIPEARIT---EAKRTMDETLNSSVQ 388
QY 55 GOKTPVNTIIPTSOPFPASILKQITL-----PKNITLSLQASPTOK- 96
DB 389 SIRTQAPTVQPVVSPNNQQLHTVLTQIVPLTVIASTDPASGTSOKFTLQAISSQP 448
QY 97 -NRIKENY 103
DB 449 MTVLKENY 456
RESULT 14
AAM38686
ID AAM38686 standard; Protein: 595 AA.
XX
XX AAM38686;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polypeptide SEQ ID NO 1831.
DE
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
XX Homo sapiens.
XX
XX

```

XX PN WO200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US34263.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX DR WPI: 2001-442253/47.
XX DR N-PSDB; AA157842.
XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX PS Example 3; SEQ ID NO 1831; 10078pp; English.
XX CC The invention relates to human nucleic acids (AA157798-AA161369) and
XX CC the encoded polypeptides (AA38642-AA42213) with nootropic,
XX CC immunosuppressant and cyostatic activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC localised neuropathies and central nervous system diseases, such as
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilisation of the activities such as: Immune system suppression,
XX CC Actin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX CC assays for receptor activity, arthritis and inflammation, leukaemias and
XX CC C.N.S disorders.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification.
XX SQ Sequence 595 AA;
XX
XX Query Match 13.0%; Score 73; DB 22; Length 595;
XX Best Local Similarity 28.9%; Pred. No. 22;
XX Matches 37; Conservative 11; Mismatches 50; Indels 30; Gaps 6;
XX
XX OY 1 ACPGVL--QPAIGPTGSSKQLFSLFHVQO---PSGNGKQVTTISHSSTLTIOKC 54
XX DB 338 ACPSEVLRVQPTQSPYPT---QLFRVHVQPVQAVPEG--EAARTSTMQDETLLNSVQ 392
XX OY 55 GQKTMPEVNTIIPTSOPPPASILKQITL-----PKNKILSLQASPTQR- 96
XX DB 393 SIRTIOAPTOVPVPSBRNQQLHTVTLQTVPLTTVASTDPASGTSQKFIQLQAISSQP 452
XX OY 97 -NRIRENV 103
XX DB 453 MIVLKENV 460
XX
XX RESULT 15
XX ID AA38685
XX AC AA38685 standard; Protein; 619 AA.
XX XX
XX DT 22-OCT-2001 (first entry)

```

```

XX DE Human polypeptide SEQ ID NO 1830.
XX KW Human: nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
XX KW peripheral nervous system; neuropathy; central nervous system; CNS;
XX KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX KW Leukemia.
XX OS Homo sapiens.
XX PN WO200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US34263.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX DR WPI: 2001-442253/47.
XX DR N-PSDB; AA157841.
XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX PS Example 3; SEQ ID NO 1830; 10078pp; English.
XX CC The invention relates to human nucleic acids (AA157798-AA161369) and
XX CC the encoded polypeptides (AA38642-AA42213) with nootropic,
XX CC immunosuppressant and cyostatic activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC localised neuropathies and central nervous system diseases, such as
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilisation of the activities such as: Immune system suppression,
XX CC Actin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX CC assays for receptor activity, arthritis and inflammation, leukaemias and
XX CC C.N.S disorders.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification.
XX SQ Sequence 619 AA;
XX
XX Query Match 13.0%; Score 73; DB 22; Length 619;
XX Best Local Similarity 28.9%; Pred. No. 22;
XX Matches 37; Conservative 11; Mismatches 50; Indels 30; Gaps 6;
XX
XX OY 1 ACPGVL--QPAIGPTGSSKQLFSLFHVQO---PSGNGKQVTTISHSSTLTIOKC 54
XX DB 362 ACPSEVLRVQPTQSPYPT---QLFRVHVQPVQAVPEG--EAARTSTMQDETLLNSVQ 416
XX OY 55 GQKTMPEVNTIIPTSOPPPASILKQITL-----PKNKILSLQASPTQR- 96
XX DB 417 SIRTIOAPTOVPVPSBRNQQLHTVTLQTVPLTTVASTDPASGTSQKFIQLQAISSQP 476
XX OY 97 -NRIRENV 103

```

Thu Feb 20 16:36:44 2003

us-09-763-909-2_copy_443_552.rag

Page 9

Db 477 MFTVLKENV 484

Search completed: February 16, 2003, 21:55:10
Job time : 10.4325 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:59:29 ; Search time 2.82655 Seconds
(without alignments)
1145.042 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_552

Perfect score: 562
Sequence: 1 AQPGLVLSQAPGIPGTSSSK.....ASPTQKNRIKENTVSCFRDE 110

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCITUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	562	100.0	801	1 US-08-725-012-2	Sequence 2, Appl1
2	92.5	16.5	737	1 US-08-168-582-16	Sequence 16, Appl1
3	92.5	16.5	737	1 US-08-646-715-16	Sequence 16, Appl1
4	72.5	12.9	412	4 US-09-134-001C-3773	Sequence 3773, Ap
5	67.5	12.0	2442	4 US-09-514-247A-10	Sequence 10, Appl1
6	66	11.7	552	1 US-08-588-821-2	Sequence 2, Appl1
7	66	11.7	552	2 US-08-915-214-2	Sequence 2, Appl1
8	66	11.7	552	2 US-09-005-532-2	Sequence 2, Appl1
9	66	11.7	552	2 US-08-701-582D-14	Sequence 14, Appl1
10	66	11.7	552	4 US-09-096-776B-9	Sequence 9, Appl1
11	66	11.7	552	4 US-09-312-748-4	Sequence 4, Appl1
12	65.5	11.7	628	4 US-09-134-001C-4461	Sequence 4461, Ap
13	65.5	11.7	628	2 US-07-952-853-22	Sequence 22, Appl1
14	65.5	11.7	628	2 US-08-914-848-22	Sequence 22, Appl1
15	65.5	11.7	1201	4 US-09-098-901-2	Sequence 2, Appl1
16	65	11.6	448	3 US-08-476-509B-2	Sequence 2, Appl1
17	65	11.6	453	4 US-08-821-278A-2	Sequence 2, Appl1
18	65	11.6	486	3 US-08-348-518C-2	Sequence 2, Appl1
19	65	11.6	896	1 US-08-095-737-2	Sequence 2, Appl1
20	65	11.6	896	1 US-08-480-145-2	Sequence 2, Appl1
21	65	11.6	896	1 US-08-477-389-2	Sequence 2, Appl1
22	64.5	11.5	903	2 US-08-853-310-2	Sequence 2, Appl1
23	64.5	11.5	2441	1 US-08-194-468-2	Sequence 2, Appl1
24	64.5	11.5	2441	1 US-08-961-739-2	Sequence 2, Appl1
25	64.5	11.5	2441	4 US-09-514-247A-8	Sequence 8, Appl1
26	64	11.4	666	2 US-08-083-590A-17	Sequence 17, Appl1
27	64	11.4	666	2 US-08-346-128-36	Sequence 36, Appl1

28	64	11.4	666	3 US-08-532-384-17	Sequence 17, Appl1
29	64	11.4	1394	4 US-09-213-053-2	Sequence 2, Appl1
30	63.5	11.3	398	4 US-09-213-053-3	Sequence 3, Appl1
31	63.5	11.3	553	4 US-08-997-251-4	Sequence 4, Appl1
32	63.5	11.3	3079	5 PCT-US94-00198-4	Sequence 4, Appl1
33	63	11.2	295	1 US-08-317-223-1	Sequence 1, Appl1
34	63	11.2	295	3 US-09-059-849A-1	Sequence 1, Appl1
35	63	11.2	295	3 US-09-213-632-1	Sequence 1, Appl1
36	63	11.2	295	3 PCT-US95-12675-1	Sequence 1, Appl1
37	63	11.2	344	4 US-09-147-236-11	Sequence 11, Appl1
38	63	11.2	770	1 US-08-445-135-2	Sequence 2, Appl1
39	62.5	11.1	726	4 US-09-126-980-2	Sequence 2, Appl1
40	62.5	11.1	726	4 US-09-476-482-2	Sequence 2, Appl1
41	62.5	11.1	726	4 US-09-517-605-6	Sequence 6, Appl1
42	62.5	11.1	1863	1 US-08-598-591-2	Sequence 2, Appl1
43	62.5	11.1	1863	1 US-08-798-691-2	Sequence 2, Appl1
44	62.5	11.1	1863	1 US-08-798-691-4	Sequence 4, Appl1
45	62.5	11.1	1863	1 US-08-798-691-6	Sequence 6, Appl1

ALIGNMENTS

```
RESULT 1
US-08-725-012-2
; Sequence 2, Application US/08725012
; Patent No. 5710025
GENERAL INFORMATION:
APPLICANT: Dikstein, Rivka
APPLICANT: Tjian, Robert
TITLE OF INVENTION: B-Cell Specific Transcription Factor
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentlin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/725,012
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: B97-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 801 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-725-012-2
Query Match 100.0%; Score 562; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 9e-63;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AQPGLVLSQAPGIPGTSSSKQLFSLFHYVQPPSGGNGKQVTTTSHSSTLTQCKGCKTNP 60
DB 443 AQPGLVLSQAPGIPGTSSSKQLFSLFHYVQPPSGGNGKQVTTTSHSSTLTQCKGCKTNP 502
OY 61 VNTTTPSQPPASLTQKLTTPGKNTLSLQASPTQKNRIKENTVSCFRDE 110
|||||
```

```
Db 503 VNTIIPTSOPFPASILKQITLPNGKILSLQASPTQKRIKENTSCFRDE 552
RESULT 2
US-08-188-582-16
; Sequence 16, Application US/08188582
; Patent No. 5534410
GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,582
; FILING DATE: 28-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 737 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-188-582-16

Query Match 16.5%; Score 92.5; DB 1; Length 737;
Best Local Similarity 27.3%; Pred. No. 0.0046;
Matches 33; Conservative 19; Mismatches 50; Indels 19; Gaps 5;

OY 2 QPGPVLS--OPAGIPNGSSSKOLFSLFHVVOQPSGNGEKQVTTISHSSTLTIOKCGOKTM 59
Db 370 QP-EVLSLTQPTGVGVGKQGPPLVIOQPPKPGALIRPPQVTLTQTPVVALNQ----- 422
OY 60 PVTIIP-----SOPFPASILKQITLPNGKIL---SLQASPTQKRIKENTSCFRD 109
Db 423 PHNRIMLTTPQOIQIOLNPLQPVVPRVAVLPGTALSAVSAQAAAKKRLKEPGGGSFRD 482
OY 110 E 110
Db 483 D 483

RESULT 3
US-08-646-715-16
; Sequence 16, Application US/08646715
; Patent No. 5637686
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,715
; FILING DATE: 09-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/188,582
; FILING DATE: 28-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 737 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-646-715-16

Query Match 16.5%; Score 92.5; DB 1; Length 737;
Best Local Similarity 27.3%; Pred. No. 0.0046;
Matches 33; Conservative 19; Mismatches 50; Indels 19; Gaps 5;

OY 2 QPGPVLS--OPAGIPNGSSSKOLFSLFHVVOQPSGNGEKQVTTISHSSTLTIOKCGOKTM 59
Db 370 QP-EVLSLTQPTGVGVGKQGPPLVIOQPPKPGALIRPPQVTLTQTPVVALNQ----- 422
OY 60 PVTIIP-----SOPFPASILKQITLPNGKIL---SLQASPTQKRIKENTSCFRD 109
Db 423 PHNRIMLTTPQOIQIOLNPLQPVVPRVAVLPGTALSAVSAQAAAKKRLKEPGGGSFRD 482
OY 110 E 110
Db 483 D 483

RESULT 4
US-09-134-001C-3773
; Sequence 3773, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
```



```

: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 3773
: LENGTH: 412
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3773

Query Match
Best Local Similarity 25.0%; Pred. No. 0.69;
Matches 28; Conservative 17; Mismatches 50; Indels 17; Gaps 4;

QY 6 VLSQPGIPGTSSSKQLFSLFHVVOQPSGNE--KQVTTIS---HSTLTIOKCGOKTM 59
Db 170 VIAGAGGLPVGTGKTLMLSGCIDSPVAGIEVMKGVTEAIIHFHSPTSEKAKDKVI 229

QY 60 PVNTI-----IPTSGPPASILKQITLPKNKILSLQASPTOKRIKENYV 104
Db 230 ELTHILAEVGPVKILHLVPEIIOKQI---NKVYHPRVTMTSTRMMMRIS 277

RESULT 5
US-09-514-247A-10
: Sequence 10, Application US/09514247A
: Patent No. 6365361
: GENERAL INFORMATION:
: APPLICANT: TANABE SEIYAKU CO. LTD.
: APPLICANT: TANIGUCHI, Tomoyasu
: TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND ANTAGONIST TO PPA
: FILE REFERENCE: TANIGUCHI-6
: CURRENT APPLICATION NUMBER: US/09/514,247A
: CURRENT FILING DATE: 2000-02-28
: PRIOR APPLICATION NUMBER: PCT/JP98/03734
: PRIOR FILING DATE: 1998-08-24
: PRIOR APPLICATION NUMBER: JP231084/1997
: PRIOR FILING DATE: 1997-08-27
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: Patentin version 3.0
: SEQ ID NO 10
: LENGTH: 2442
: TYPE: PRT
: ORGANISM: human
US-09-514-247A-10

Query Match
Best Local Similarity 12.0%; Score 67.5; DB 4; Length 2442;
Matches 30; Conservative 19; Mismatches 41; Indels 45; Gaps 5;

QY 2 QPGVYLSPPAIP-----TGSSSKQLFSLFHVVOQPSGNE--KQV----- 40
Db 1910 QPSVSMSPAPFPVARTQPTVTSTGKPTQVPAPPPAPPAVAARIEAEAOQO 1969

QY 41 -----TTSISS-----TTLIOKCGOKTMPTVNTIIPTSOPPPASILKO 78
Db 1970 QHLRYVNTNNMSPGRTGCTGPGSOMAPVSLNVRPNQVSGPVMSMPGQMOQAPLPOQ 2029

QY 79 ITLPG--NKILSLQA 91
Db 2030 QPMGGLPRPVISMQA 2044

RESULT 6
US-08-588-821-2
: Sequence 2, Application US/08588821
: Patent No. 5712097
: GENERAL INFORMATION:
: APPLICANT: Kern, Scott E.
: APPLICANT: Hahn, Stephan A.
```

```

: TITLE OF INVENTION: NOVEL TUMOR SUPPRESSOR GENE, DPC4
: NUMBER OF SEQUENCES: 91
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 4225 Executive Square, Suite 1400
: CITY: La Jolla
: STATE: CA
: COUNTRY: USA
: ZIP: 92037
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/588,821
: FILING DATE: 19-JAN-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Halle, Lisa A.
: REGISTRATION NUMBER: 38,347
: REFERENCE/DOCKET NUMBER: 07265/079001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619/678-5070
: TELEFAX: 619/678-5099
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 552 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-588-821-2

Query Match
Best Local Similarity 11.7%; Score 66; DB 1; Length 552;
Matches 22; Conservative 7; Mismatches 28; Indels 10; Gaps 3;

QY 8 SQPAGIPGTSSSKQLFSLFHVVOQPSGNEK-----QVTTISSTLTIOKCGOKTM 62
Db 223 SQPAGILGSHSE---GLQIASGPPQOQNGFTGQATYHNHSTTT--WTGSTRATYT 277

QY 63 TITPTSQ 69
Db 278 PNLPHHQ 284

RESULT 7
US-08-915-214-2
: Sequence 2, Application US/08915214
: Patent No. 5814457
: GENERAL INFORMATION:
: APPLICANT: Kern, Scott E.
: APPLICANT: Hahn, Stephan A.
: TITLE OF INVENTION: NOVEL TUMOR SUPPRESSOR GENE, DPC4
: NUMBER OF SEQUENCES: 91
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 4225 Executive Square, Suite 1400
: CITY: La Jolla
: STATE: CA
: COUNTRY: USA
: ZIP: 92037
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/915,214
: FILING DATE: 20-AUG-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/588,821
```


APPLICANT: Kawabata, Masahiro
TITLE OF INVENTION: SMAD6 AND USES THEREOF
FILE REFERENCE: L0461/7038
CURRENT APPLICATION NUMBER: US/09/096,776B
CURRENT FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: US 60/049,990
PRIOR FILING DATE: 1997-06-13
PRIOR APPLICATION NUMBER: US 60/053,040
PRIOR FILING DATE: 1997-07-18
PRIOR APPLICATION NUMBER: US 60/066,173
PRIOR FILING DATE: 1997-11-18
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn version 3.0
SEQ ID NO 9
LENGTH: 552
TYPE: PRF
ORGANISM: Homo sapiens
US-09-096-776B-9

Query Match 11.7%; Score 66; DB 4; Length 552;
Best Local Similarity 32.8%; Pred. No. 7.1;
Matches 22; Conservative 7; Mismatches 28; Indels 10; Gaps 3;

QY 8 SQPAGIPTGSSKQLFLFHVVOOPSGNKK-----QVTTISHSTLTLCCKGCKTAPVN 62
DB 223 SQPASILGSHSE---GLDQIASGPQPQOQNGFTGQPATYHNHSTTT--WTGSRITAPYT 277

QY 63 TIPTSQ 69
DB 278 PNLPHHQ 284

RESULT 11
US-09-312-748-4
Sequence 4, Application US/09312748
Patent No. 6423491
GENERAL INFORMATION:
APPLICANT: HOME, JAMES R.
APPLICANT: AALTONEN, LAURI A.
TITLE OF INVENTION: A GENE FOR JUVENILE POLYPOSIS
FILE REFERENCE: IOWA:021
CURRENT APPLICATION NUMBER: US/09/312,748
CURRENT FILING DATE: 1999-05-13
EARLIER APPLICATION NUMBER: 60/085,312
EARLIER FILING DATE: 1998-05-13
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 552
TYPE: PRF
ORGANISM: Homo sapiens
FEATURE:
US-09-312-748-4

Query Match 11.7%; Score 66; DB 4; Length 552;
Best Local Similarity 32.8%; Pred. No. 7.1;
Matches 22; Conservative 7; Mismatches 28; Indels 10; Gaps 3;

QY 8 SQPAGIPTGSSKQLFLFHVVOOPSGNKK-----QVTTISHSTLTLCCKGCKTAPVN 62
DB 223 SQPASILGSHSE---GLDQIASGPQPQOQNGFTGQPATYHNHSTTT--WTGSRITAPYT 277

QY 63 TIPTSQ 69
DB 278 PNLPHHQ 284

RESULT 12
US-09-134-001C-4461
Sequence 4461, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4461
LENGTH: 451
TYPE: PRF
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4461

Query Match 11.7%; Score 65.5; DB 4; Length 451;
Best Local Similarity 24.6%; Pred. No. 6.1;
Matches 31; Conservative 20; Mismatches 40; Indels 35; Gaps 5;

QY 6 VLSOPAGIPTGSSKQLFLFHVVOOPSGNKKQVTTISHSTLT-----TIOKC 54
DB 75 VKSRATSIKRAKTSKQALIS-----PKTSSTKTTTAKKSTYVAKTTTTRKQPTIRKS 127

QY 55 G-----OKTMEVN-----TIPTSQPPASILKQITLPKNTLSIQASPTOKNRI 99
DB 128 STTSRATKMTPTSVKRTSKATVTPSPSKAKISRTKQQT--KSHTSVAKNTTQLSKT 185

QY 100 KENYTS 105
DB 186 KSPSTS 191

RESULT 13
US-07-952-853-22
Sequence 22, Application US/07952853
Patent No. 5863783
GENERAL INFORMATION:
APPLICANT: Van Heuvel, Margaretha
APPLICANT: Bakhuys, Janna G.
APPLICANT: Coustel, Yves
APPLICANT: Harder, Abraham
APPLICANT: De Graaff, Leendert H.
APPLICANT: Filippi, Michel J. A.
APPLICANT: Van Der Veen, Peter
APPLICANT: Visser, Jacob
APPLICANT: Andreoli, Peter M.
TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA
TITLE OF INVENTION: MOLECULES
TITLE OF INVENTION: ENCODING ARABIAN-DEGRADING ENZYMES OF FUNGAL
TITLE OF INVENTION: ORIGIN
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (ERO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/952,853
FILING DATE: 19921125
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murashige, Kate H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 246152003500
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-813-5600

TELEFAX: 415-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 628 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-952-853-22

Query Match 11.7%; Score 65.5; DB 2; Length 628;
Best Local Similarity 26.2%; Pred. No. 9.9;
Matches 22; Conservative 13; Mismatches 30; Indels 19; Gaps 3;

Db 123 GIPVDS--EFHSSFWIKDYSGDITVRLVGNVTGTEGSTITHTST-----ADNFTQ 174
QY 61 VNTIPTSOPPASILKQITLPGN 84
Db 175 ASVKFPTKAPDGNVLYELTVDS 198

RESULT 14
US-08-914-848-22
Sequence 22, Application US/08914848
Patent No. 5989887
GENERAL INFORMATION:
APPLICANT: Van Heuvel, Margaretha
APPLICANT: Bakhuys, Janna G.
APPLICANT: Coutel, Yves
APPLICANT: Harder, Abraham
APPLICANT: De Graaff, Leendert H.
APPLICANT: Elipphl, Michel J. A.
APPLICANT: Van Der Veen, Peter
APPLICANT: Visser, Jacob
TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA
TITLE OF INVENTION: MOLECULES ENCODING ARABINAN-DIGRADING ENZYMES OF FUNGAL
TITLE OF INVENTION: ORIGIN
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/914,848
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/952,853
FILING DATE: 25-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Murashige, Kate H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 246152003500
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-813-5600
TELEFAX: 415-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 628 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-914-848-22

Query Match 11.7%; Score 65.5; DB 2; Length 628;
Best Local Similarity 26.2%; Pred. No. 9.9;
Matches 22; Conservative 13; Mismatches 30; Indels 19; Gaps 3;

Db 123 GIPVDS--EFHSSFWIKDYSGDITVRLVGNVTGTEGSTITHTST-----ADNFTQ 174
QY 61 VNTIPTSOPPASILKQITLPGN 84
Db 175 ASVKFPTKAPDGNVLYELTVDS 198

RESULT 15
US-09-098-901-2
Sequence 2, Application US/09098901B
Patent No. 6218144
GENERAL INFORMATION:
APPLICANT: Scott, Mathew
APPLICANT: Sisson, John C.
TITLE OF INVENTION: Costal2 Genes and their Uses
FILE REFERENCE: SUN-65P
CURRENT APPLICATION NUMBER: US/09/098,901B
EARLIER FILING DATE: 1998-06-17
EARLIER APPLICATION NUMBER: 60/051,347
NUMBER OF SEQ ID NOS: 11
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1201
TYPE: PRT
ORGANISM: D. Melanogaster
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)-(1201)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-098-901-2

Query Match 11.7%; Score 65.5; DB 4; Length 1201;
Best Local Similarity 22.0%; Pred. No. 26;
Matches 26; Conservative 18; Mismatches 43; Indels 31; Gaps 3;

QY 6 VLSOPAGIPPGS-----SKQLPSLFHVVOO-----PSGNEKQ----- 39
Db 1081 LIRAPKPMPTGSLADRYKDEKGRSRIIFAKFVLRYSAAAGSGSTAEESTALIDS 1140
QY 40 -----VTTTSHSSTLTIOKCGCKTMEVNTIPTSOPPASILKQITLPGKITLSQAS 92
Db 1141 TTTATVTTSTTTGAVGKVKALVSFRPEQLRLMPAPATATKVTROKKNITIDAS 1198

Search completed: February 16, 2003, 22:03:01
Job time : 4.82655 secs

;
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO

APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5708
LENGTH: 215
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5708

Query Match 12.2%; Score 68.5; DB 9; Length 215;
Best Local Similarity 25.2%; Pred. No. 3.2;
Matches 26; Conservative 19; Mismatches 35; Indels 23; Gaps 5;
DB 19 SKOLESLFHVQPSGNGKQVTTTSHSST-----LTIQKCGQRTMPVNTIIPTSQF-70
DB 68 TSHISLV-----QPLKGPPEAVSTIRATSSRYERKHGKAMAECSST---GMICPSFAFE 120
QY 71 ----PPAS-----ILKQITLPGNKIISLQASPTQKNRIKENYVS 105
DB 121 VNSPATSDSLAKRACAPDSRVASVAPSPKPTALSTVSA 163

RESULT 3
US-09-893-519A-14
Sequence 14, Application US/09893519A
Publication No. US20030027243A1
GENERAL INFORMATION:
APPLICANT: ANADYS PHARMACEUTICALS, INC.
APPLICANT: THOMPSON, Craig
APPLICANT: MOORE, Jeffrey
APPLICANT: BUDEMAN, Ed T.
APPLICANT: BRADLEY, John
APPLICANT: DESTIVA, Thamara
APPLICANT: HARRIS, Sandra
APPLICANT: KOMARNITSKY, Svetlana
APPLICANT: MENDILLO, Marc
APPLICANT: MOORE, Daniel
APPLICANT: MCCOY, Melissa
APPLICANT: SANDERSON, Karen
APPLICANT: HAQ, Tariq
APPLICANT: ZHU, Shuhao
APPLICANT: LONG, Fan
TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
FILE REFERENCE: 0342/16548-US2
CURRENT APPLICATION NUMBER: US/09/893,519A
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/215,164
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: US 60/224,457
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 146
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 1023
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:

NAME/KEY: misc_feature
OTHER INFORMATION: Corresponds to SEQ ID NO: 87
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Human Genbank/CAA72189
DATABASE ENTRY DATE: 1997-06-25
RELEVANT RESIDUES: (1)..(1023)
US-09-893-519A-14

Query Match 12.1%; Score 68; DB 9; Length 1023;
Best Local Similarity 29.8%; Pred. No. 27;
Matches 28; Conservative 11; Mismatches 45; Indels 10; Gaps 4;
DB 5 PVLSOPAGIPTGSSKQLESLFHVQPSGNGKQVTTTSHSSTLTIQKCG-----QKTP 60
DB 432 PRLPQPNQNTNQNQLPPGVNLVRSNG-----QLMTIQQLAQMOQAHAQPTTMA 487
QY 61 VNTIIPTSQFPPASITLQITLPGNKIISLQASPT 94
DB 488 PRATPTSA-PPVQT-STVQAPGPTIARQVTF 519

RESULT 4
US-10-109-886-10
Sequence 10, Application US/10109886
Patent No. US20020119499A1
GENERAL INFORMATION:
APPLICANT: TANABE SEIYAKU CO. LTD.
APPLICANT: TANIGUCHI, Tomoyasu
APPLICANT: MIZUKAMI, Junko
TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND
FILE REFERENCE: TANIGUCHI-6
CURRENT APPLICATION NUMBER: US/10/109,886
CURRENT FILING DATE: 2002-04-01
PRIOR APPLICATION NUMBER: 09/514,247
PRIOR FILING DATE: 2000-02-28
PRIOR APPLICATION NUMBER: PCT/JP98/03734
PRIOR FILING DATE: 1998-08-24
PRIOR APPLICATION NUMBER: JP231084/1997
PRIOR FILING DATE: 1997-08-27
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn version 3.0
SEQ ID NO 10
LENGTH: 2442
TYPE: PRT
ORGANISM: human
US-10-109-886-10

Query Match 12.0%; Score 67.5; DB 12; Length 2442;
Best Local Similarity 22.2%; Pred. No. 94;
Matches 30; Conservative 19; Mismatches 41; Indels 45; Gaps 5;
QY 2 QPGPVLSOPAGIP-----TGSSSKQLESLFHVQPSGNGE---KQV----- 40
DB 1910 QPSPVSMSPAGPSPVARQTPTTVSTGKPTSOVAPAPPPAPPAVEARQTERAQOQ 1969
QY 41 -----TTSHS-----TLTIQKCGQRTMPVNTIIPTSQFPPASILKQ 78
DB 1970 QHLYRVNINNSMPGRGTGTPGSOQAPVSLNVRBNVSGPVPMPGMOQAAPLPOQ 2029
QY 79 ITLPG--NKIISLQ 91
DB 2030 QPMPGLRPVISMQA 2044

RESULT 5
US-09-815-242-10928
Sequence 10928, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10928
LENGTH: 375
TYPE: PRF
ORGANISM: Enterococcus faecalis
US-09-815-242-10928

Query Match 11.9%; Score 67; DB 10; Length 375;
Best Local Similarity 27.8%; Pred. No. 9.6;

Matches 20; Conservative 19; Mismatches 29; Indels 4; Gaps 3;
DB 71 PPSILKQITLP 82
207 -PGTTLKIEIP 217

RESULT 6
US-09-815-242-12792
Sequence 12792, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12792
LENGTH: 748
TYPE: PRF
ORGANISM: Staphylococcus aureus
US-09-815-242-12792

Query Match 11.9%; Score 67; DB 10; Length 748;
Best Local Similarity 25.2%; Pred. No. 23;

Matches 27; Conservative 23; Mismatches 35; Indels 22; Gaps 6;
DB 8 SOPAGIPGSSSKQLF-----SLFHYVQOPSGGNEKQVTTISHST--LTTQKCGQKTM 60
225 SGP-----SRKKRVFAENSSNNIVNHQADQOOLTEQTHNSVESENTEAGEVT-N 277
DB 61 VNTIIPTSQPPASILKQITLPGNKILSLQASPTQKNRIKENVTSCF 107
278 VSYVV-----PPLTLLNO---PAKQATSKAEVQKGOVLENTLKDF 316

RESULT 7
US-09-815-242-12327
Sequence 12327, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12327
LENGTH: 792
TYPE: PRF
ORGANISM: Staphylococcus aureus
US-09-815-242-12327

Query Match 11.9%; Score 67; DB 10; Length 792;
Best Local Similarity 25.2%; Pred. No. 25;

Matches 27; Conservative 23; Mismatches 35; Indels 22; Gaps 6;
DB 8 SOPAGIPGSSSKQLF-----SLFHYVQOPSGGNEKQVTTISHST--LTTQKCGQKTM 60
269 SGP-----SRKKRVFAENSSNNIVNHQADQOOLTEQTHNSVESENTEAGEVT-N 321

OY 61 VNTIPTSOFPASILKQITLPGNKILSLQASPTOKRIKENYTSOF 107
DB 322 VSTIV-----PPLTLNO---PAKOKATSKAEVQKRGQVLENTLKDF 360

RESULT 8
US-09-923-922-9
; Sequence 9, Application US/09923922
; Patent No. US20020106732A1
; GENERAL INFORMATION:
; APPLICANT: Miyazono, Kohel
; APPLICANT: Kawabata, Masahiro
; TITLE OF INVENTION: SMAD6 AND USES THEREOF
; FILE REFERENCE: L0461/7120
; CURRENT APPLICATION NUMBER: US/09/923,922
; CURRENT FILING DATE: 2001-08-07
; PRIOR APPLICATION NUMBER: 09/096,776
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: US 60/049,990
; PRIOR FILING DATE: 1997-06-13
; PRIOR APPLICATION NUMBER: US 60/053,040
; PRIOR FILING DATE: 1997-07-18
; PRIOR APPLICATION NUMBER: US 60/066,173
; PRIOR FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 552
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-923-922-9

Query Match 11.7%; Score 66; DB 10; Length 552;
Best Local Similarity 32.8%; Pred. No. 20;
Matches 22; Conservative 7; Mismatches 28; Indels 10; Gaps 3;

OY 8 SOPAGIPTSSSKQLFSLFHVQOPSGNGEK-----QVTTISSTLTIOGCGOKTIPV 62
DB 223 SQPASLIGSHS-----GLQIASGPGQOQNGFTGQPRATYHNNSTTT--WTGSRAPRT 277

OY 63 TITPSQ 69
DB 278 PMLPHNQ 284

RESULT 9
US-09-964-899-33
; Sequence 33, Application US/09964899
; Patent No. US2002017446A1
; GENERAL INFORMATION:
; APPLICANT: Cohen, Dalia et al.
; TITLE OF INVENTION: Identification of Genes Involved in
; FILE REFERENCE: 4-31612 A
; CURRENT APPLICATION NUMBER: US/09/964,899
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/236,893
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/298,309
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 33
; LENGTH: 1249
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-964-899-33

Query Match 11.7%; Score 66; DB 9; Length 1249;
Best Local Similarity 21.3%; Pred. No. 58;
Matches 27; Conservative 22; Mismatches 54; Indels 24; Gaps 5;

OY 1 AOPGVLG---PAGIPTSSSKQLFSLFHVQOPSGNGEKQVTTISSTLTIOGCGOK 57

DB 892 ANPGQTESSKFFPGIOAKSHRNITKLL---EDEPRSRDETPICCTIAHMODSLAKRC--- 945
OY 58 TMEVNTIPTSOFPASILKQITLPG-----NKILSLQASPTOKNRI-----KENV 103
DB 946 -ICVSNIVRSLSFVPGDADMSKHPGLVLLTKLILHHEHPKRRAPQYKEKEDBDKG 1004

OY 104 TSCPRDE 110
DB 1005 VACSKDE 1011

RESULT 10
US-09-815-242-14110
; Sequence 14110, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14110
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(343)
; OTHER INFORMATION: Xaa - Any Amino Acid
US-09-815-242-14110

Query Match 11.7%; Score 65.5; DB 10; Length 343;
Best Local Similarity 24.1%; Pred. No. 12;
Matches 26; Conservative 20; Mismatches 41; Indels 21; Gaps 4;

OY 7 LSPAGI-----PTGSSSKQLFSLFHVQOPSGN-----EKVTTISSTLTIOGCGO 56
DB 26 LHPVAGQIYGVIGASGKSTLIRCVNLERPREGSYMGQDELTTLSSEGLTKARR--- 82

OY 57 KTMPVNTIPTSOFPASILKQITLPGNKILSLQASPTOKRIKENY 104
DB 83 ---QIGMI-----FQHFNLISRTVFGNVALPLELNTFKKEETKRKYV 122

RESULT 11
US-09-964-899-25
; Sequence 25, Application US/09964899
; Patent No. US2002017446A1


```

1 RESULT 14
2 US-09-729-674-176
3 ; Sequence 176: Application US/09729674
4 Patent No. US20010039335A1
5 ; GENERAL INFORMATION:
6 APPLICANT: Jacobs, Kenneth
7 APPLICANT: McCoy, John M.
8 APPLICANT: Lavaille, Edward R.
9 APPLICANT: Collins-Racie, Lisa A.
10 APPLICANT: Evans, Cheryl
11 APPLICANT: Merberg, David
12 APPLICANT: Treacy, Maurice
13 APPLICANT: Agostino, Michael J.
14 APPLICANT: Steininger II, Robert J.
15 APPLICANT: Spaulding, Vikki
16 APPLICANT: Wong, Gordon G.
17 APPLICANT: Clark, Hilary
18 APPLICANT: Fechtel, Kim
19 APPLICANT: Genetics Institute, Inc.
20 TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
21 FILE REFERENCE: 6035-64X
22 CURRENT APPLICATION NUMBER: US/09/729,674
23 CURRENT FILING DATE: 2000-12-04
24 PRIOR APPLICATION NUMBER: 09/539,330
25 PRIOR FILING DATE: 2000-03-30
26 NUMBER OF SEQ ID NOS: 283
27 SOFTWARE: PatentIn Ver. 2.0
28 SEQ ID NO 176
29 LENGTH: 832
30 TYPE: PRT
31 ORGANISM: Homo sapiens
32 FEATURE:

```

NAME/KEY: UNSURE
LOCATION: (12)
NAME/KEY: UNSURE
LOCATION: (449)
US-09-729-674-176

Query Match

11.4%; Score 64; DB 10; Length 832;

Best Local Similarity 26.2%; Pred. No. 57;

Matches 32; Conservative 14; Mismatches 46; Indels 30; Gaps 5;

QY 2 QGPGVLT-SOPACIPGTSSSKQLEFLFHVYQPGSGNEKQVTTISHSTLTIOKCGQKMP 60
DB 206 QNGCTVLPSEAGLATASCPITVSSVAASQGLCVTNTRTSVRKQ---LFCVPRKTS 261
QY 61 VNTIT-----PTSGPPASITKQITLPGNKILSQASPTQKN 97
DB 262 PAVIVSVSTCSSLPVSSAPITSGCAPPT-FLPAS-TSQADLSQKMFSAVPPTKE 319
QY 98 RI 99
DB 320 KV 321

QY 98 RI 99
DB 320 KV 321

RESULT 15

US-09-864-761-43246

Sequence 43246, Application US/09864761

Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

FILE REFERENCE: Aecm1ca-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/774,203

PRIOR FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 43246
LENGTH: 432
TYPE: PRT
ORGANISM: Homo sapiens

FEATURE: MAP TO AC005859.1

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.99

OTHER INFORMATION: EST HUMAN HIT: AM867810.1, EVALUATE 5.00e-13

OTHER INFORMATION: SWISSPROT HIT: Q28888, EVALUATE 4.00e-14

US-09-864-761-43246

Query Match

11.3%; Score 63.5; DB 10; Length 432;

Best Local Similarity 24.0%; Pred. No. 28;

Matches 31; Conservative 14; Mismatches 31; Indels 53; Gaps 6;

QY 10 PAGIPGTSSSKQLEFLFHVYQPGSGNEKQVTTISHSTLTIO-----KCGQ 56
DB 59 PGGIPNTNTNLT-----TINHPDISPASFHRDLHVEIDFRCNC 99
QY 57 KTMPV-----NTIIPTSQPPASI-----LKOITLPGNKIL-----SLQASPTQKN 99
DB 100 VPIPIGSKNNMCIKRLQIKRPSFSGLTYSYLDGNGQLLEIFQGLPSPQLLSLEANNI 159
QY 100 -----KENVT 104
DB 160 FSIRKENLIT 168

Search completed: February 16, 2003, 22:02:37
Job time: 5.16702 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2003, 22:03:10 ; Search time 92 Seconds
(without alignments)
8526.947 Million cell updates/sec

Title: US-09-763-909-1

Perfect score: 2558
Sequence: 1 gggaccctgtgaccaaagt.....cttattacttacctat 2558

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgnt2_6/ptodata/2/1na/5A_COMB.seq.*
2: /cgnt2_6/ptodata/2/1na/5B_COMB.seq.*
3: /cgnt2_6/ptodata/2/1na/6A_COMB.seq.*
4: /cgnt2_6/ptodata/2/1na/6B_COMB.seq.*
5: /cgnt2_6/ptodata/2/1na/PCTUS_COMB.seq.*
6: /cgnt2_6/ptodata/2/1na/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2534	99.1	2556	1 US-08-725-012-1	Sequence 1, Appl
2	338.2	13.2	3603	1 US-08-188-582-15	Sequence 15, Appl
3	338.2	13.2	3603	1 US-08-646-715-15	Sequence 15, Appl
4	142.2	5.6	4615	1 US-08-188-582-1	Sequence 1, Appl
5	142.2	5.6	4615	1 US-08-646-715-1	Sequence 1, Appl
6	57.6	2.3	7218	1 US-08-232-463-14	Sequence 14, Appl
7	41	1.6	2169	4 US-09-434-408-3	Sequence 3, Appl
8	37	1.4	2223	1 US-08-257-073-4	Sequence 4, Appl
9	36	1.4	998	1 US-09-122-4008-5	Sequence 5, Appl
10	35.4	1.4	397	3 US-09-253-691-3	Sequence 3, Appl
11	35.4	1.4	5163	3 US-08-700-651-1	Sequence 1, Appl
12	35.4	1.4	5163	3 US-08-928-361B-4	Sequence 4, Appl
13	35.4	1.4	5318	3 US-08-700-651-2	Sequence 2, Appl
14	35.4	1.4	5318	3 US-08-928-361B-3	Sequence 3, Appl
15	35.2	1.4	2030	4 US-09-512-342-1	Sequence 1, Appl
16	35	1.4	543	6 5273901-6	Patent No. 5273901
17	35	1.4	2190	4 US-09-625-188-19	Sequence 19, Appl
18	34.8	1.4	19307	3 US-08-836-022A-10	Sequence 10, Appl
19	34.8	1.4	19307	4 US-09-427-048A-10	Sequence 10, Appl
20	34.6	1.4	1778	1 US-08-416-870C-3	Sequence 3, Appl
21	34.4	1.3	387	4 US-09-615-192A-253	Sequence 253, App
22	34.4	1.3	1916	4 US-09-484-970B-6	Sequence 6, Appl
23	34.4	1.3	21338	4 US-08-961-527-20	Sequence 20, Appl
24	34.2	1.3	376	2 US-08-623-906A-18	Sequence 18, Appl
25	34.2	1.3	1008	4 US-09-527-345-1	Sequence 1, Appl
26	34.2	1.3	1170	4 US-08-973-005A-9	Sequence 9, Appl
27	34.2	1.3	2338	4 US-08-845-258-13	Sequence 13, Appl

C 28	34.2	1.3	2338	4 US-08-990-571-13	Sequence 13, Appl
C 29	34.2	1.3	2338	4 US-08-723-142A-13	Sequence 13, Appl
C 30	34.2	1.3	2338	4 US-09-528-784A-13	Sequence 13, Appl
31	34	1.3	3666	2 US-08-682-517-13	Sequence 2, Appl
32	34	1.3	3666	2 US-08-682-517-14	Sequence 14, Appl
33	34	1.3	4197	2 US-08-682-517-7	Sequence 7, Appl
34	34	1.3	4197	2 US-08-682-517-8	Sequence 8, Appl
35	34	1.3	7705	2 US-08-687-080-115	Sequence 115, Appl
36	33.8	1.3	1695	1 US-08-361-920-20	Sequence 20, Appl
37	33.8	1.3	1695	1 US-08-479-939-20	Sequence 20, Appl
38	33.8	1.3	1695	1 US-08-483-432-20	Sequence 20, Appl
39	33.8	1.3	2088	4 US-09-351-414-3	Sequence 3, Appl
C 40	33.2	1.3	98844	4 US-09-791-211-10	Sequence 8, Appl
41	33	1.3	21040	4 US-08-961-527-55	Sequence 55, Appl
42	32.8	1.3	1545	4 US-08-900-117A-2	Sequence 2, Appl
43	32.8	1.3	2088	1 US-08-331-394-1	Sequence 1, Appl
44	32.8	1.3	2088	1 US-08-250-858-1	Sequence 1, Appl
45	32.8	1.3	2088	1 US-08-446-915-1	Sequence 1, Appl

ALIGNMENTS

```
RESULT 1
US-08-725-012-1
: Sequence 1, Application US/08725012
: Patent No. 5710025
:
: GENERAL INFORMATION:
: APPLICANT: Dikstein, Rivka
: APPLICANT: Tjian, Robert
: TITLE OF INVENTION: B-Cell Specific Transcription Factor
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
: STREET: 268 BUSH STREET, SUITE 3200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94104
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/725,012
: FILING DATE:
:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: OSMAN, RICHARD A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: B97-005
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 343-4341
: TELEFAX: (415) 343-4342
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2556 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..2403
:
: US-08-725-012-1
:
: Query Match 99.1%; Score 2534; DB 1; Length 2556;
: Best Local Similarity 99.9%; Pred. No. 0;
: Matches 2556; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
```

Dh 1 GGGACCCTGGTGAACAAAGTGGCTCCGGTCAGGCGCCCTCTTAAGTACAGAGCGGCCCT 60
Qy 61 AGGCTGCTCTCTCTCTGATAGTGCCTGGTAAAGCCCCCAACACACAGACATCTCAATTT 120
Dh 61 AGGCTGCTCTCTCTCTGATAGTGCCTGGTAAAGCCCCCAACACACAGACATCTCAATTT 120
Qy 121 CTTGCTAATTTGGAGCTTCTCTCCAGAAACGTTTGTATTTAAAGTAAACAGTGGTCCGTTG 180
Dh 121 CTTGCTAATTTGGAGCTTCTCTCCAGAAACGTTTGTATTTAAAGTAAACAGTGGTCCGTTG 180
Qy 181 ATGTTGGTATCTCTCTGACAACTGTAAACAAGCCGAGACACAACTAACTAACTCA 240
Dh 181 ATGTTGGTATCTCTCTGACAACTGTAAACAAGCCGAGACACAACTAACTAACTCA 240
Qy 241 AGGCCAGACATACGAGGAATCTCTCAAAACGTCAAAATCTGTACAGTCCGAACTTACG 300
Dh 241 AGGCCAGACATACGAGGAATCTCTCAAAACGTCAAAATCTGTACAGTCCGAACTTACG 300
Qy 301 TCACAAATTAATCAAGAAAGTGGCAGTGAACACCTGTTAAAAAATGGCACAATAGSAACT 360
Dh 301 TCACAAATTAATCAAGAAAGTGGCAGTGAACACCTGTTAAAAAATGGCACAATAGSAACT 360
Qy 361 AGTGTGGTAAACAATGTTCCGAAGCCCTTCTCAGTACAACTGTGGCTGTGCCAAACAGT 420
Dh 361 AGTGTGGTAAACAATGTTCCGAAGCCCTTCTCAGTACAACTGTGGCTGTGCCAAACAGT 420
Qy 421 RGTGTCACAGTTACACCTCTGGAAGCCATGTGAATCTGTACATCTGTGAAGCTTAAAG 480
Dh 421 -GTGCTACAGTTACACCTCTGGAAGCCATGTGAATCTGTACATCTGTGAAGCTTAAAG 480
Qy 481 TTTGGGAGCATCATCCACTCTCTCAAAATGAGCCCAATCTTAAAGCAGAGAACTCAGAGC 540
Dh 481 TTTGGGAGCATCATCCACTCTCTCAAAATGAGCCCAATCTTAAAGCAGAGAACTCAGAGC 540
Qy 541 TGTTCAGATTAATCTTCTTCGACAAATGTAAGAAATGTGAAGAAATGCAAGAACTTCT 600
Dh 541 TGTTCAGATTAATCTTCTTCGACAAATGTAAGAAATGTGAAGAAATGCAAGAACTTCT 600
Qy 601 TGAATGTTAAATAAATACATGATGTAGTACAGTCCCTCGAATTTGGGGCAAAATGT 660
Dh 601 TGAATGTTAAATAAATACATGATGTAGTACAGTCCCTCGAATTTGGGGCAAAATGT 660
Qy 661 GAAGAAGCTGTGGAAACATTTTGGATGTGAAGAAATGGAAGCAGAAAGATTTACTAGAA 720
Dh 661 GAAGAAGCTGTGGAAACATTTTGGATGTGAAGAAATGGAAGCAGAAAGATTTACTAGAA 720
Qy 721 ACTGTAATGTTGAACATGATCTTACCTCAGCTCAGCTGCTTCTTCTTAAAGAAAG 780
Dh 721 ACTGTAATGTTGAACATGATCTTACCTCAGCTCAGCTGCTTCTTCTTAAAGAAAG 780
Qy 781 CGTGTGCTTACGACAACTTCTGCTTAATCTCCAGAGTTCATCCAGCAATGTGTCA 840
Dh 781 CGTGTGCTTACGACAACTTCTGCTTAATCTCCAGAGTTCATCCAGCAATGTGTCA 840
Qy 841 GCAGACTTCTAGTACATGATGCTATCTACTACTACAAAGTAAACAATCTCTCTGT 900
Dh 841 GCAGACTTCTAGTACATGATGCTATCTACTACTACAAAGTAAACAATCTCTCTGT 900
Qy 901 GGTGACAACTACAGTGTCTCTCAAGCAAGTGTGAAGAAATGTAATTTGTTTCTGGAGAAC 960
Dh 901 GGTGACAACTACAGTGTCTCTCAAGCAAGTGTGAAGAAATGTAATTTGTTTCTGGAGAAC 960
Qy 961 AGCACCAGAACTGTGCTGCAAGTGTGAAGCACTGTCGTCGTCGTCGTCGTCGTCGTCG 1020
Dh 961 AGCACCAGAACTGTGCTGCAAGTGTGAAGCACTGTCGTCGTCGTCGTCGTCGTCGTCG 1020
Qy 1021 AGCTGAGTGTGACACTTCTCTGTGGCCCAACTGTCTGCAACAGAGAAACAAACAGC 1080
Dh 1021 AGCTGAGTGTGACACTTCTCTGTGGCCCAACTGTCTGCAACAGAGAAACAAACAGC 1080
Qy 1081 TGGAACTGTTTGTCTCAGACTTCAAAACCACTTGTGACATCTGTGGCAAAACACAGTAC 1140
Dh 1081 TGGAACTGTTTGTCTCAGACTTCAAAACCACTTGTGACATCTGTGGCAAAACACAGTAC 1140
Qy 1080 TGGAACTGTTTGTCTCAGACTTCAAAACCACTTGTGACATCTGTGGCAAAACACAGTAC 1139

Qy 1141 CAGGCTCTCACTGCAACCTGGAAGCCAGTGTCTCTGGAACACAGTAACTGTCTCT 1200
Dh 1140 CAGGCTCTCACTGCAACCTGGAAGCCAGTGTCTCTGGAACACAGTAACTGTCTCT 1199
Qy 1201 TCCAGCAGTAACTTTTGGAGAACTTCAAGTGTGAGCTATTTGTCTTCATCTGTGAAGC 1260
Dh 1200 TCCAGCAGTAACTTTTGGAGAACTTCAAGTGTGAGCTATTTGTCTTCATCTGTGAAGC 1259
Qy 1261 TGTGTTTCTCTCTGCTGGGACCAATCTGCAAGCCCTGTTAATGGAGCTCAGTTCAAT 1320
Dh 1260 TGTGTTTCTCTCTGCTGGGACCAATCTGCAAGCCCTGTTAATGGAGCTCAGTTCAAT 1319
Qy 1321 CAAACTTGGCCAGCGGCGCTGTCTTTCACAAACAGCTGGAGTTCACAGGCACTTC 1380
Dh 1320 CAAACTTGGCCAGCGGCGCTGTCTTTCACAAACAGCTGGAGTTCACAGGCACTTC 1379
Qy 1381 AAGCAGCACTATTTCAATGTTTCAAGTTCAGAGCTCAGAGCTTCAAGAGCAATGAAA 1440
Dh 1380 AAGCAGCACTATTTCAATGTTTCAAGTTCAGAGCTCAGAGCTTCAAGAGCAATGAAA 1439
Qy 1441 ACAAGTACCAACAATTTCAATCTCTCAACATTTGACCAATTCAGAAATGTGGACAGAGAC 1500
Dh 1440 ACAAGTACCAACAATTTCAATCTCTCAACATTTGACCAATTCAGAAATGTGGACAGAGAC 1499
Qy 1501 GATGCGAGTGAACCAATTAATCTACTAGTCACTTCTCAGCTTCTCAATCTTAAAGCA 1560
Dh 1500 GATGCGAGTGAACCAATTAATCTACTAGTCACTTCTCAGCTTCTCAATCTTAAAGCA 1559
Qy 1561 AATTAACCTGCTGCTGGAATTAATTTCTGTCACATCTCAAGATCTCTCAAGCAAGAAATA 1620
Dh 1560 AATTA -CTTGTCTGGAATTAATTTCTGTCACATCTCAAGATCTCTCAAGCAAGAAATA 1618
Qy 1621 GAATTAAGAAAGAAATGTAATCATCTCTCCAGATGAGATGACATCAATGATGTGACTT 1680
Dh 1619 GAATTAAGAAAGAAATGTAATCATCTCTCCAGATGAGATGACATCAATGATGTGACTT 1678
Qy 1681 CTATGCGAGGGGTCAACCTTAATGAGAAATGCTCTCATCTTTCAGCAAACTGTGAAT 1740
Dh 1679 CTATGCGAGGGGTCAACCTTAATGAGAAATGCTCTCATCTTTCAGCAAACTGTGAAT 1738
Qy 1741 TGGTTGACACATCTCAATGATGTAAGTGAAGCAATTTCTTAAATGGAGCTCTAC 1800
Dh 1739 TGGTTGACACATCTCAATGATGTAAGTGAAGCAATTTCTTAAATGGAGCTCTAC 1798
Qy 1801 AAAAGAGAAATCTTGAACATTTGTAAAAAGCATGATTAAGAACTTAACTGTGATGCTG 1860
Dh 1799 AAAAGAGAAATCTTGAACATTTGTAAAAAGCATGATTAAGAACTTAACTGTGATGCTG 1858
Qy 1861 TGAACCTGATCTCCCAAGCAACAGAGACGACACAGGCTTCTTGAAGAAACTGACTG 1920
Dh 1859 TGAACCTGATCTCCCAAGCAACAGAGACGACACAGGCTTCTTGAAGAAACTGACTG 1918
Qy 1921 CAATTTGCTCAGATCGAATGATCTTAACAAGCAATGTAATTAATCTGTGATGCTG 1980
Dh 1919 CAATTTGCTCAGATCGAATGATCTTAACAAGCAATGTAATTAATCTGTGATGCTG 1978
Qy 1981 ATACCAAGTCAAGCTCAAAATTTCTTGAAGAGCTGATCAATTTGAGAAACAGAGAAAG 2040
Dh 1979 ATACCAAGTCAAGCTCAAAATTTCTTGAAGAGCTGATCAATTTGAGAAACAGAGAAAG 2038
Qy 2041 ATTTGGAAGAAAGAAATTTTACTTTAAGCAGCAAGATCTTCTTAATTAAGAAATGATC 2100
Dh 2039 ATTTGGAAGAAAGAAATTTTACTTTAAGCAGCAAGATCTTCTTAATTAAGAAATGATC 2098
Qy 2101 CAAACAGCTGAGATTTAAACAGAAAGCAAAGAGTTCAGCAATTTGGAATCTTGACAGA 2160
Dh 2099 CAAACAGCTGAGATTTAAACAGAAAGCAAAGAGTTCAGCAATTTGGAATCTTGACAGA 2158
Qy 2161 TACAGCATAGAGAGCTTAATCTCAGAGCTTTCAGCTATTTGAGCAAGAGAAAGAGAC 2220
Dh 2159 TACAGCATAGAGAGCTTAATCTCAGAGCTTTCAGCTATTTGAGCAAGAGAAAGAGAC 2218

QY 2221 CACTAGAAATCTGGAAATGAGGGCTTAAAGACACCTTCTGCTTCTGGACATCCAGCC 2280
| | | | |
Db 2219 CACTAGAAATCTGGAAATGAGGGCTTAAAGACACCTTCTGCTTCTGGACATCCAGCC 2278
| | | | |
QY 2281 TGACAGCCACCAAGAGTTGCACTCTCCAAAGATACAGAAATCTGCTTCCAGGACTTGA 2340
| | | | |
Db 2279 TGACAGCCACCAAGAGTTGCACTCTCCAAAGATACAGAAATCTGCTTCCAGGACTTGA 2338
| | | | |
QY 2341 TATTTTGTATGGAACAGGAGGAGATGAAATATCTGAGCTCTATATACCGGGCCCTTC 2400
| | | | |
Db 2339 TATTTTGTATGGAACAGGAGGAGATGAAATATCTGAGCTCTATATACCGGGCCCTTC 2398
| | | | |
QY 2401 TGAAGTGAACACTCCACTCTTCATCCACATCTTGTATTTACTGCGCAAGAGACACA 2460
| | | | |
Db 2399 TGAAGTGAACACTCCACTCTTCATCCACATCTTGTATTTACTGCGCAAGAGACACA 2458
| | | | |
QY 2461 AAGCATTTGTGACCTGCTCTGAAATTTCAATTTCTGGAATAAACCACCACTGAAAGAG 2520
| | | | |
Db 2459 AAGCATTTGTGACCTGCTCTGAAATTTCAATTTCTGGAATAAACCACCACTGAAAGAG 2518
| | | | |
QY 2521 CATTTGTACGATTTAGAACTTATTTAGCTTACCTTACT 2558
| | | | |
Db 2519 CATTTGTACGATTTAGAACTTATTTAGCTTACCTTACT 2556
| | | | |

RESULT 2
US-08-188-582-15
; Sequence 15, Application US/08188582
; Patent No. 5534410
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comal, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS.
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TATS AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188, 582
; FILING DATE: 28-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AUT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 272299
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3603 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:

NAME/KEY: CDS
LOCATION: 1..2214
US-08-188-582-15

Query Match 13.2%; Score 338.2; DB 1; Length 3603;
Best Local Similarity 63.9%; Pred. No. 6.3e-92;
Matches 536; Conservative 0; Mismatches 288; Indels 15; Gaps 1;

QY 1586 TCTGTCACCTTCAAGATATCTCCATCCAGCAAAATATGAAATTAAGAAATGTAACATCATG 1645
| | | | |
Db 1377 TCTCTCGGCACAGAGAGCTCTGTCACAGAAATTAATCAAGAGCCTGGGGAGCTTC 1436
| | | | |
QY 1646 CTTCGAGATGAGATGATGATCATATGATGTGATCTTATGCGAGGGGTCAACTTAATGA 1705
| | | | |
Db 1437 GTTGGGAGCATGATGATCATTAATGATGTGATCATATGCTGGAGTAACTTTGTCA 1496
| | | | |
QY 1706 AGAAATAGCTTCACTTTAGCAACAACCTGTAATTTGGTGGCACATCATCACTATG 1765
| | | | |
Db 1497 AGAAATAGCAAAATTAATGACGACAGACTTAATTTGGTGGCACATCATCACTATG 1556
| | | | |
QY 1766 TAAAGATGACCAATTTCTTTTATTTGAGCTCTACAAAAGAGATCTTAGACATTTGTAA 1825
| | | | |
Db 1557 TAAAGATGACCAATTTCTTTCTCTCCAAAGCGCTTTCAGAGAAATTAATTAAGTAA 1616
| | | | |
QY 1826 AAGCATGACATTTAGAGAACTTAACTGTGATGCTGATGATCTGATCTCCCAAGACACA 1885
| | | | |
Db 1617 AAAACATGATTAAGAGAAATTAATCAATCCAGATAGTAAATTTATCATGATCCACGCA 1676
| | | | |
QY 1886 GGAACGACTACGAGGCTTTTGAAAAACTGACATGCTGATGCTGATGATGATGATG 1945
| | | | |
Db 1677 ACMAAGGCTACAGAAATCTTTGAGAGAAATTAATCAAAACAGCTCAGAGAAAGCTTTC 1736
| | | | |
QY 1946 TTACAGAGCAATGTAATTAATTAATCATCTGTTGATGATGATGATGATGATGATGAT 2005
| | | | |
Db 1737 TTACAGAGCAATGTAATTAATTAATCATCTGTTGATGATGATGATGATGATGATGAT 1796
| | | | |
QY 2006 TGAAGAGCTGATCAATTTGAGAGACAGAGAAAGATTTGGAAGAAAGAAATGTTACT 2065
| | | | |
Db 1797 TGAAGAGCTGATCAATTTGAGAGAAAGATTTGGAAGAAAGATTTGGAAGAAATGTTACT 1856
| | | | |
QY 2066 TGAAGAGCTGATCAATTTGAGAGAAAGATTTGGAAGAAAGATTTGGAAGAAATGTTACT 2125
| | | | |
Db 1857 GAGGCGACGAAAGCTGCTGCTCAAGACAGAAAGATGATGATGATGATGATGATGATG 1916
| | | | |
QY 2126 AGCCAAAGCTTACAGCAATTTGCAAGACTGACAGATGATGATGATGATGATGATGATG 2185
| | | | |
Db 1917 GCAAGAGATGACAGCAAGAGAACTGACAGAAATGATGATGATGATGATGATGATGATG 1976
| | | | |
QY 2186 AGCTCTTGCACTATTGACCAAGAGAAAGAGACACTAGA-----ATC 2230
| | | | |
Db 1977 AGCAGTACAGGAGATCGGGCCAGGAAAGAGAAAGATGATGATGATGATGATGATGATG 2036
| | | | |
QY 2231 TCGAATTTGAGGCTTAAAGAGACACCTTCTTCTGAGGATCCAGGCTTACAGCCAC 2290
| | | | |
Db 2037 AGGAGCAGAGGGGTGGGGCCCGGCTCAATGCTCCAGGACCTGAGGTGTGGAAACCC 2096
| | | | |
QY 2291 CAACAGTTGCATTCGTCACCAAGATGACAGATGATGATGATGATGATGATGATGATG 2350
| | | | |
Db 2097 CAGACATTCACGCGCAAGAGATCACGGGCTCAACCTCAAGGAGCTCATATTTGTTT 2156
| | | | |
QY 2351 GGAACAGAGAGGAGATGAAATATCTGAGCTTATACCTGAGCCCTTCTGAAAGTGA 2409
| | | | |
Db 2157 AGAAATGAAAGCTGAGAGCAAGCAATTCAGCTGCTGCTTCAAAAGCATTCCTTAAAGTGA 2215
| | | | |

RESULT 3
US-08-646-715-15
; Sequence 15, Application US/08646715
; Patent No. 5637686
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comal, Lucio
; APPLICANT: Dynlacht, Brian D.

```

APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/646,715
FILING DATE: 09-MAY-1996
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/188,582
FILING DATE: 28-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 3603 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2214
US-08-646-715-15
Query Match 13.2%; Score 338.2; DB 1; Length 3603;
Best Local Similarity 63.9%; Pred. No. 6.3e-92;
Matches 536; Conservative 0; Mismatches 288; Indels 15; Gaps 1;

```

```

Db 1677 ACAAAGCTACAGATCTTTAGAGAAATATCAAGAAACAGCTACGAGAAAGAACTTTTC 1736
Qy 1946 TTACAGGCAAGTGAATAATATACATCTGTGTAGTATACAGCTACAGCTCAATTTCT 2005
Db 1737 TTACAGGATGACAGATATATGAGAGGCGAGTGTGGGCGACAGCTCAAGTTTTC 1796
Qy 2006 TGAAGAGTGTATATTTGGAAGCAAGAGAAAGATTTGGAGAAAGAAATGTTACT 2065
Db 1797 TGAACAGCTGTATCAATTCGAAAGAGAGAGAGAGATGACGAGCGGAGATCTGAT 1856
Qy 2066 TAAGCAGCCAGAGCTGTCTATTAAGAAAGATCCAGAACAGCTGATTAAGCAGAA 2125
Db 1857 GAGGCGCAAGAGTCTGTCTGTAAGAGAAAGATCCAGAAAGTAAAGCTGAAACGAA 1916
Qy 2126 AGCCAAAGAGTTACAGCAATTTGAATCTGACAGATACAGATAGAGACCTAATCTAC 2185
Db 1917 GCGAAAGAGATGCGAGCAAGAGAACTGGGCAATATGACAGGCGGAGCCCAACCTCAC 1976
Qy 2186 AGCTCTGACGCTATTGGACCAAGAGAAAGAGACCACTAGA-----ATC 2230
Db 1977 AGCACTAGCAGCGATCGGGCCAGAGAAAGAGAGAGTGTACTGTCCGGGCGGCTC 2036
Qy 2231 TGAATGAGGGCTTAAAGACACCTTCCTTGGGACATCCAGCCTGACAGCCAC 2290
Db 2037 AGGAGCAGAGGGGTCCGGCCCGCTCAGTGTCTCCAGCACTCGGCTGGAACCC 2096
Qy 2291 CAACAGTTGCATCTCCAGAAATCAGAGATCTGCTCAGAGGACTTGATATTTGTAT 2350
Db 2097 CAGACAGTTACCGCGCAAGAAATCAGCGGGCTCACTCAGGAGCTCATATTTGTTT 2156
Qy 2351 GGAACAGAAACGGAGATGAAGTATTTCTGAGCTCTTACTGGCTTTGAGATGAC 2409
Db 2157 AGAAATGAACGTGAGCAAGCAAGCTTCAGCTGCTCTACAAAGCAATTCCTTAAGTAC 2215

RESULT 4
US-08-188-582-1
Sequence 1, Application US/08188582
Patent No. 5534410
GENERAL INFORMATION:
APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dynlacht, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,582
FILING DATE: 28-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989

```

TELEFAX: (415) 398-3249
 TELE: 910 277299
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4615 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: CDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 538..3300
 US-08-188-582-1

Query Match 5.6%; Score 142.2; DB 1; Length 4615;
 Best Local Similarity 50.7%; Pred. No. 1.7e-32;
 Matches 395; Conservative 0; Mismatches 378; Indels 6; Gaps 2;

QY 1653 GATGAGATGACATCATGATGCTTCTATGCGAGGGGTCACCTTAATGAGAAAAT 1712
 DB 2548 GGTGACGATGATATCAACGATGTTCCGCCATGGAGGTGTTAACTGGCGAGAGTCG 2607
 QY 1713 GCCTGCATCTAGCAACAACCTGTAATGTTGGCACATCATTCATCATTAAGAT 1772
 DB 2608 CAGCGAATTCCTGGCTGTA---CGAAAAACATCGGACGAGATTGCAATCTCGAAAGAT 2664
 QY 1773 GAACCATTTCTTTTATTGAGCTCTACAAAAGAAATCTTAGACATTGTTAAAGCAT 1832
 DB 2665 GAGGTTTTCTTAATCTCCCTCGCTGCAAGTGAATACGGGCAATTAATTCTCGAGGCG 2724
 QY 1833 GACATTACAGAACTTAATCTGATGCTGTGAATCTGATCTCCCAACACAGAGAGA 1892
 DB 2725 GGACTGATGAGCCGCTCCAGAGATGTGCGCTTCTGATATGCGACGCTGTCAGAGCGC 2784
 QY 1893 CTACGAGGCTCTTGTGAAAACCTGCAATTTGCTCAGATCGAATGATCTACTTAACAAG 1952
 DB 2785 CTGAAGAACATCTGTTGAGAAAGTTGGCTGTGATAGCGAGACCCGATTGATGATCAATCAAG 2844
 QY 1953 GCAGTGAATTTACATCTCTGTAGTATACACAGTCAACGCTCAAAATTTCTTGAAAAG 2012
 DB 2845 TTGATCCACGCTATGAGCCGCCCAAGGATGTGGCGGTCAAGATCAATTTCTCGAGGAG 2904
 QY 2013 CTGATCAATTGGAGACAGAGAAAGATTGGAAAGAAAGAAATGTTACTTAAGCA 2072
 DB 2905 CTGGCAAGGCCGAGACAGAGACAGAGAACTGAGAGCTGATGATCTGCTGCGGGCA 2964
 QY 2073 GCCAAGAGTCTGTCTAATAAAGAAAGATCCAGACAGCTGAGATTAAAGCAAGAAACCAAA 2132
 DB 2965 GCCAAGTACGCTGAGGCTGAGAGATCCCGAGCAGGCCAAGATGAAGCGGAGGCCAAG 3024
 QY 2133 GAGTTACAGCAATTGGAATCTGCACAGATACAGCATTAAGACGCTAATCTCACAGCTCTT 2192
 DB 3025 GAGATGCAAGCGCGCGAATGAGAGATGTGCTCAACAGAGATCCAAATCTGACGGCGCTG 3084
 QY 2193 GCAGCTATTGACCAAGAGAGAGAGACCACTAGATCTGGAATTTAGAGGCTTTAAAGAC 2252
 DB 3085 CAGGGGATGAGACCTCGGAAAAAGCTGAAGCTGAGCGGCGAAGCAGTCAAGTTCGGGAGCG 3144
 QY 2253 AACCTTTTGGCTCTG---GACATCCAGCGCTGACAGCCACCAACAGTTGATGATGCTCA 2309
 DB 3145 GGTTCAGATGGCGGCGGAGCTGTAAGCAGCTCGGATCTGCGCGACAGCTTTACGGCTT 3204
 QY 2310 AGAATCAGAGAAATCTGCTCAGGACTTGAATTTTATGATGAGACAGAAAGGGAGATG 2369
 DB 3205 CGCATAAAAAGTGTACCTGCGGAGACATGCTCTTCTACATGAGACAAAGAGGGAGTTT 3264
 QY 2370 AAGATTTCAGAGCTCTTACTGCGCTTCTGAAAGTACCACTCACTCTTCCATCCA 2428
 DB 3265 TGTGCAATTCATGCTGTGTAAGACATACCTCAAGTATGATGCTGTTGCCATCA 3323

RESULT 5
 US-08-646-715-1

Sequence 1, Application US/08646715
 Patent No. 5637686
 GENERAL INFORMATION:
 APPLICANT: Tjian, Robert
 APPLICANT: Comal, Lucio
 APPLICANT: Dynlacht, Brian D.
 APPLICANT: Hoey, Timothy
 APPLICANT: Ruppert, Siegfried
 APPLICANT: Tanese, Naoko
 APPLICANT: Wang, Edith
 APPLICANT: Weitzel, Robert O.J.
 TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
 TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAPS AND METHODS OF USE
 NUMBER OF SEQUENCES: 36
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/646,715
 FILING DATE: 09-MAY-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/188,582
 FILING DATE: 28-JAN-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Osman, Richard A.
 REGISTRATION NUMBER: 36,627
 REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249

QY 1653 GATGAGATGACATCATGATGCTTCTATGCGAGGGGTCACCTTAATGAGAAAAT 1712
 DB 2548 GGTGACGATGATATCAACGATGTTCCGCCATGGAGGTGTTAACTGGCGAGAGTCG 2607
 QY 1713 GCCTGCATCTAGCAACAACCTGTAATGTTGGCACATCATTCATCATTAAGAT 1772
 DB 2608 CAGCGAATTCCTGGCTGTA---CGAAAAACATCGGACGAGATTGCAATCTCGAAAGAT 2664
 QY 1773 GAACCATTTCTTTTATTGAGCTCTACAAAAGAAATCTTAGACATTGTTAAAGCAT 1832
 DB 2665 GAGGTTTTCTTAATCTCCCTCGCTGCAAGTGAATACGGGCAATTAATTCTCGAGGCG 2724
 QY 1833 GACATTACAGAACTTAATCTGATGCTGTGAATCTGATCTCCCAACACAGAGAGA 1892
 DB 2725 GGACTGATGAGCCGCTCCAGAGATGTGCGCTTCTGATATGCGACGCTGTCAGAGCGC 2784

Query Match 5.6%; Score 142.2; DB 1; Length 4615;
 Best Local Similarity 50.7%; Pred. No. 1.7e-32;
 Matches 395; Conservative 0; Mismatches 378; Indels 6; Gaps 2;

QY 1653 GATGAGATGACATCATGATGCTTCTATGCGAGGGGTCACCTTAATGAGAAAAT 1712
 DB 2548 GGTGACGATGATATCAACGATGTTCCGCCATGGAGGTGTTAACTGGCGAGAGTCG 2607
 QY 1713 GCCTGCATCTAGCAACAACCTGTAATGTTGGCACATCATTCATCATTAAGAT 1772
 DB 2608 CAGCGAATTCCTGGCTGTA---CGAAAAACATCGGACGAGATTGCAATCTCGAAAGAT 2664
 QY 1773 GAACCATTTCTTTTATTGAGCTCTACAAAAGAAATCTTAGACATTGTTAAAGCAT 1832
 DB 2665 GAGGTTTTCTTAATCTCCCTCGCTGCAAGTGAATACGGGCAATTAATTCTCGAGGCG 2724
 QY 1833 GACATTACAGAACTTAATCTGATGCTGTGAATCTGATCTCCCAACACAGAGAGA 1892
 DB 2725 GGACTGATGAGCCGCTCCAGAGATGTGCGCTTCTGATATGCGACGCTGTCAGAGCGC 2784

QY 1893 CTACGAGGCGCTTCAGAAAAACTGACGCAATTGCTCAGCATGCAATCTACTTACAG 1952
DB 2785 CTAGAGAACATCGTTGGAAGTTGGCTGTATATGCGGAGACCGCATTTGATCATCAG 2844
QY 1953 GCAAGTGAATAATTCATCTGTGTAGTATCAGAGTCAAGCTCAATTTCTTGAAGA 2012
DB 2845 TTGATTCACAGCTATGAGCCCGCAAGAGATGTGGCGGTGAGATCAAGTTCTCGAGAG 2904
QY 2013 CTGATCATATGAGAAAGCAGAAAGGATTTGAGAAAGAAATGTTACTTAAAGCA 2072
DB 2905 CTGGACAAAGGCGAGCAAGAGACAGAGAACTGAGAGGTGATGCTGCTGGGCA 2964
QY 2073 GCCAAGAGCTCTTAATAAGAAAGATCCAGAACGCTGAGATTAAGAAAGCCAAA 2132
DB 2965 GCCAAGTCAAGCTGAGAGGTGGAAGATCCGAGACAGCCAAAGTGAAGGCGAGCCAA 3024
QY 2133 GAGTACAGCAATTTGAACTTGCACAGATACAGCATAGAGACGCTTAATTCACAGCT 2192
DB 3025 GAGATGCAAGCGCGCAAAATGAGAGAGTTCGTCACAGATGCAATTCGAGCGGCTG 3084
QY 2193 GCAAGCTATTTGACCAAGAAAGAGACCACTAGAAATCTGGAATTTGAGGCTTAAAG 2252
DB 3085 CAGCGCATTTGAGCTGGAAGAAAGCTGAACTGAGCGCAAAACAGTCACTGCGGAGCG 3144
QY 2253 AACCTTCTTCTCTG--GACATCCAGCGCTGACAGCACAAACGTTGCAATGTCGA 2309
DB 3145 GGTTCAGATGCGCGGAGTGTAGCAGCTCGGATCTCGCGGACAGCTTACGGCTT 3204
QY 2310 AGAATCAGCAAGAACTGCTCAGGAGCTGTAATTTTGTATGAAACAGAAAGGAGATG 2369
DB 3205 CGCATTAACGTGTGAACCTGCGCAGCATGCTCTTTCATACATGGAAGCAAGGAGATTC 3264
QY 2370 AAGTATTTCTGAGCTATATCTGCGGCTTCTGAAAGTCACTCCACTTTCCATCA 2428
DB 3265 TGTGCAAGTTCATGCTGTTCAAGACATACCTCAAGTATGCTGCTGCTGCCATCA 3323

RESULT 6
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)836-9300
TELEFAX: (703)863-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZ9PL-F15
US-08-232-463-14

Query Match 2.3%; Score 57.6; DB 1; Length 7218;
Best Local Similarity 3.8%; Pred. No. 9.2e-07;
Matches 15; Conservative 226; Mismatches 155; Indels 0; Gaps 0;

QY 1885 AGCAAGCACTAGAGGCGCTCTAGAAAAAGCATGCAATTCACAGCATGCAATGACTA 1944
DB 1436 ACRR 1377
QY 1945 CTTACAAAGCAAGTGAATAATTCATCTCTGTAGTATACAGTCAAGCTCAATTTTC 2004
DB 1376 RRR 1317
QY 2005 TTGAAAAGCTGATCAATTTGACAGACAGCAAGAAAGATTGGAAGAAAGAAATGTTAC 2064
DB 1316 RRR 1257
QY 2065 TTAAGCAGCAGCAGAGCGTTCTTAATAAGAAAGTCCAGAAAGCTGAGATTTAAAGCAG 2124
DB 1256 RRR 1197
QY 2125 AAGCAAGAGTACCAATTTGCACTGACAGATACAGATACAGATACAGATACAGATAC 2184
DB 1196 RRR 1137
QY 2185 CAGCTTTCAGCTATTTGACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2244
DB 1136 RRR 1077
QY 2245 TAAAGACACCTTCTTGTCTTGGGACATCCAGCC 2280
DB 1076 RRR 1041

RESULT 7
US-09-434-408-3
Sequence 3, Application US/09434408
Patent No. 6440697
GENERAL INFORMATION:
APPLICANT: Venezia, Domenick
APPLICANT: Grossmann, Angelika
TITLE OF INVENTION: RING FINGER PROTEIN ZAPO3
FILE REFERENCE: 98-41
CURRENT APPLICATION NUMBER: US/09/434,408
CURRENT FILING DATE: 1999-11-04
EARLIER APPLICATION NUMBER: US 60/108,258
EARLIER FILING DATE: 1998-11-12
NUMBER OF SEQ ID NOS: 23
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 2169
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Degenerate polynucleotide sequence of zapop3
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(2169)
OTHER INFORMATION: n = A,T,C or G
US-09-434-408-3

QY 2057 AATGTTACTTAAGGACGACCAAGTCTGCTTAATTAAGAAAGATCCAGACAGTACATTT 2116
DB 749 AACTGCTTTAAAGACTGAAATAATTTCTTAATAATATTTTGTAAATAACTGAAAAA 690
QY 2117 AAAGCAGAAAGCCAAAGAGTTACAGCAATTTGAACTTGACAGCA 2160
DB 689 AAAGCAGAAAGTAATTTTCTTAAGTAATTTAAAACTGAAAAA 646

RESULT 10

US-09-253-691-3
; Sequence 3, Application US/09253691
; Patent No. 6124100
; GENERAL INFORMATION:
; APPLICANT: Dong Kyu JIN
; TITLE OF INVENTION: Diagnostic Method and Kit for Neuropsychiatric Diseases
; FILE REFERENCE: 1942/36
; CURRENT APPLICATION NUMBER: US/09/253,691
; EARLIER FILING DATE: 1999-02-22
; EARLIER FILING DATE: 1996-02-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: WordPerfect 6.1/Windows
; SEQ ID NO 3
; LENGTH: 397
; TYPE: DNA
; ORGANISM: human
US-09-253-691-3

Query Match 1.4%; Score 35.4; DB 3; Length 397;
Best Local Similarity 47.2%; Pred. No. 0.76;
Matches 108; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 1008 CAGTGGGAGCAAAAGCTGAGTTGTGACACTTCTGTGGGCCCAACTGTCACACAG 1067
DB 125 CAAAACAGACAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 184
QY 1068 GAGAACAAACAGCTGGAACGTGTTGCTTCACTCACTCACTCACTCACTCACTCACT 1127
DB 185 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 244
QY 1128 CAACACAGTACAGCAGCTGCTCACTGCAACCTGAAAGCAAGTGTCTGGAACAGCAG 1187
DB 245 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 304
QY 1188 TAACACTGTCCTTCCAGCAGTAATTTTGGAGAACTTCAGTGACG 1236
DB 305 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 353

RESULT 11

US-08-700-651-1
; Sequence 1, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAs AND RNAs
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 5163
; TYPE: DNA

; ORGANISM: Cryptosporidium parvum
US-08-700-651-1

Query Match 1.4%; Score 35.4; DB 3; Length 5163;
Best Local Similarity 45.5%; Pred. No. 3.9;
Matches 126; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 863 ATTGCTACCTGTACTACACAGTAACACTTCTCTGTGTGACACTACAGTGTCTCA 922
DB 737 ACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 796
QY 923 AGCAGCTGTAAGCAATTAATTTGTTTGTGAGCAGCAGCAGCAGCAGCAGCAGCAGC 982
DB 797 ACCCAACTACACCAACCACTACCAACCAACCAACCAACCAACCAACCAACCAACCA 856
QY 983 CAACCTTGAACCCACTTGTGTCACAGTGGAGCAAAAGCTGGAGTTGTGACACTTCA 1042
DB 857 ACAACTCAACCAACCAACCACTACACCAACCAACCAACCAACCAACCAACCAACCA 916
QY 1043 TCTGTGGCCCAACTGCTGCAACAGAGAAACCAACAGCTGGAACCTGTTGCTTACAG 1102
DB 917 AAACCAACCAACCAACCACTACCAACCAACCAACCAACCAACCAACCAACCAACCA 976
QY 1103 TCAAAACCACTTGTGACATCTGTGCAAAACCACTGA 1139
DB 977 ACAACCAACCAACCAACCACTACCAACCAAG 1013

RESULT 12

US-08-928-361B-4
; Sequence 4, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIRKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: VERNY, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5163 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-928-361B-4

Query Match 1.4%; Score 35.4; DB 3; Length 5163;
Best Local Similarity 45.5%; Pred. No. 3.9;
Matches 126; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 863 ATTGCTACCTTACTACAGTACAGTACCTCTCTGCTGGAGACAGTACAGTCTCTCA 922
DB 736 ACTTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 795

QY 923 AGCCAGTGTGAAGTCAATTAATGTTCTGTGAGCAACAGCACCAGACCTGTGCTAGT 982
DB 796 ACCAAGTACTACACCAACCACTACACCAACCAACCAACCAACCAACCACTACAC 855

QY 983 CAATTTGAAACCACTGCTGTGCTGCTGAGGAGCAAAAGCTGAGTTGTGACACTTAT 1042
DB 856 ACACTACACCAACCACTACACCAACCAACCAACCAACCAACCACTACCAAG 915

QY 1043 TCTGTGGGCCCACTGTGCAACAGAGAACAGTGTGAGTGTGTTGCTTACACT 1102
DB 916 AAACCAACCAACCACTGTGATCTGTGCAACAGTGA 1139

QY 1103 TCAAAACCACTGTGATCTGTGCAACAGTGA 1139
DB 976 ACAACCAACCAACCACTGTGCAACAGTGA 1012

RESULT 13
US-08-700-651-2
Sequence 2, Application US/08700651B
Patent No. 6015882
GENERAL INFORMATION:
APPLICANT: PETERSEN, CAROLYN
APPLICANT: LEBCH, JAMES
APPLICANT: NELSON, RICHARD, C.
APPLICANT: GUY, JIRI
TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
FILE REFERENCE: 480.19-4(HV)
CURRENT APPLICATION NUMBER: US/08/700.651B
CURRENT FILING DATE: 1997-08-14
EARLIER APPLICATION NUMBER: 08/415.751
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 2
LENGTH: 5318
TYPE: DNA
ORGANISM: Cryptosporidium parvum
US-08-700-651-2

Query Match 1.4%; Score 35.4; DB 3; Length 5318;
Best Local Similarity 45.5%; Pred. No. 4;
Matches 126; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 863 ATTGCTACCTTACTACAGTACAGTACCTCTCTGCTGGAGACAGTACAGTCTCTCA 922
DB 737 ACTTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 796

QY 923 AGCCAGTGTGAAGTCAATTAATGTTCTGTGAGCAACAGCACCAGACCTGTGCTAGT 982
DB 797 ACCAAGTACTACACCAACCACTACACCAACCAACCAACCAACCACTACCAAG 856

QY 983 CAATTTGAAACCACTGTGCTGTGCTGAGGAGCAAAAGCTGAGTTGTGACACTTAT 1042
DB 857 ACACTTACACCAACCACTGTGATCTGTGCAACAGTGA 916

QY 1043 TCTGTGGGCCCACTGTGCAACAGAGAACAGTGTGAGTGTGTTGCTTACACT 1102
DB 917 AAACCAACCAACCACTGTGATCTGTGCAACAGTGA 976

QY 1103 TCAAAACCACTGTGATCTGTGCAACAGTGA 1139
DB 976 ACAACCAACCAACCACTGTGCAACAGTGA 1012

DB 977 ACAACCAACCAACCACTGTGCAACAGTGA 1013

RESULT 14
US-08-928-361B-3
Sequence 3, Application US/08928361B
Patent No. 6071518
GENERAL INFORMATION:
APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
TITLE OF INVENTION: SPECIES INFECTIONS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: PETERS, VERNY, JONES & BIKSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928.361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026.062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Verry, Hans
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5318 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-928-361B-3

Query Match 1.4%; Score 35.4; DB 3; Length 5318;
Best Local Similarity 45.5%; Pred. No. 4;
Matches 126; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 863 ATTGCTACCTTACTACAGTACAGTACCTCTCTGCTGGAGACAGTACAGTCTCTCA 922
DB 736 ACTTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 795

QY 923 AGCCAGTGTGAAGTCAATTAATGTTCTGTGAGCAACAGCACCAGACCTGTGCTAGT 982
DB 796 ACCAAGTACTACACCAACCACTACACCAACCAACCAACCAACCACTACCAAG 855

QY 983 CAATTTGAAACCACTGTGCTGTGCTGAGGAGCAAAAGCTGAGTTGTGACACTTAT 1042
DB 856 ACACTTACACCAACCACTGTGATCTGTGCAACAGTGA 915

QY 1043 TCTGTGGGCCCACTGTGCAACAGAGAACAGTGTGAGTGTGTTGCTTACACT 1102
DB 916 AAACCAACCAACCACTGTGATCTGTGCAACAGTGA 975

QY 1103 TCAAAACCACTGTGATCTGTGCAACAGTGA 1139
DB 976 ACAACCAACCAACCACTGTGCAACAGTGA 1012

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2003, 22:04:35 ; Search time 126 Seconds
(without alignments)
10340.079 Million cell updates/sec

Title: US-09-763-909-1

Perfect score: 2558
Sequence: 1 gggaccctgtgacccaagt.....cttataactcttacctat 2558

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 424239 seqs, 254661826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	135.8	5.3	2307	9	US-09-893-519A-87
2	38.6	1.5	5101	10	US-09-070-927A-516
3	37.8	1.5	438	10	US-09-960-352-11462
4	37.8	1.5	442	10	US-09-960-352-11227
5	36.8	1.4	579	10	US-09-864-761-9944
6	36.6	1.4	174424	10	US-09-967-768A-314
7	36	1.4	486	10	US-09-864-761-10113
8	36	1.4	510	10	US-09-864-761-18737
9	36	1.4	591	10	US-09-864-761-12567
10	36	1.4	998	10	US-09-816-894-5
11	36	1.4	2391	10	US-09-815-242-3827
12	36	1.4	2397	10	US-09-815-242-6667
13	35.8	1.4	1668	10	US-09-974-300-101
14	35.8	1.4	2534	10	US-09-764-878-403
15	35.6	1.4	640681	9	US-09-790-988-1
16	35.4	1.4	30350	9	US-10-118-328-3
17	34.8	1.4	1225	9	US-09-822-846-198
18	34.6	1.4	294	10	US-09-864-761-28655
19	34.6	1.4	561	10	US-09-864-761-12075

20	34.4	1.3	520	10	US-09-917-800A-1015	Sequence 1015, Ap
21	34.4	1.3	1137	9	US-09-815-242-9348	Sequence 9348, Ap
22	34.4	1.3	1278	10	US-09-764-868-280	Sequence 280, App
23	34.4	1.3	10249	12	US-10-002-600-65	Sequence 65, App
24	34.4	1.3	10249	12	US-10-044-090-5	Sequence 5, App1
25	34.4	1.3	57130	10	US-09-835-081-3	Sequence 3, App1
26	34.2	1.3	418	10	US-09-960-352-7989	Sequence 7989, Ap
27	34.2	1.3	1008	9	US-09-922-469-1	Sequence 1, App1
28	34.2	1.3	1008	10	US-09-922-480-1	Sequence 1, App1
29	34.2	1.3	1008	10	US-09-923-236-1	Sequence 1, App1
30	34.2	1.3	1872	10	US-09-815-242-4585	Sequence 4585, Ap
31	34.2	1.3	1878	10	US-09-815-242-8606	Sequence 8606, Ap
32	34.2	1.3	2338	9	US-09-286-488-13	Sequence 13, App1
33	34.2	1.3	2338	10	US-09-737-178-13	Sequence 13, App1
34	34.2	1.3	11854	10	US-09-070-927A-306	Sequence 306, App
35	34	1.3	385	10	US-09-960-352-1739	Sequence 1739, Ap
36	34	1.3	519	10	US-09-878-574-4292	Sequence 4292, Ap
37	34	1.3	3666	10	US-09-137-531-13	Sequence 13, App1
38	34	1.3	3666	10	US-09-137-531-14	Sequence 14, App1
39	34	1.3	4197	10	US-09-137-531-7	Sequence 7, App1
40	34	1.3	4197	10	US-09-137-531-8	Sequence 8, App1
41	34	1.3	13865	10	US-09-070-927A-200	Sequence 200, App
42	33.8	1.3	1930	12	US-10-003-152-15	Sequence 15, App1
43	33.8	1.3	2056	12	US-10-003-152-23	Sequence 23, App1
44	33.8	1.3	2056	10	US-09-809-790-3	Sequence 3, App1
45	33.8	1.3	2088	10	US-09-809-617-3	Sequence 3, App1

ALIGNMENTS

RESULT 1
US-09-893-519A-87
Sequence 87, Application US/09893519A
Publication No. US20030027243A1

GENERAL INFORMATION:
APPLICANT: ANADYS PHARMACEUTICALS, INC.
APPLICANT: THOMPSON, Craig
APPLICANT: MOORE, Jeffrey
APPLICANT: BURMAN, Ed T.
APPLICANT: BRADLEY, John
APPLICANT: DESTIVA, Thamarra
APPLICANT: HARRIS, Sandra
APPLICANT: KOMARNITSKY, Svetlana
APPLICANT: MENDILLO, Marc
APPLICANT: MOORE, Daniel
APPLICANT: MCCOY, Melissa
APPLICANT: SANDERSON, Karen
APPLICANT: HAO, Tariq
APPLICANT: ZHU, Shuhao
APPLICANT: LONG, Fan

TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
FILE REFERENCE: 0342/IG548-US2
CURRENT APPLICATION NUMBER: US/09/893,519A
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/215,164
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: US 60/224,457
PRIOR FILING DATE: 2000-06-10
NUMBER OF SEQ ID NOS: 146
SOFTWARE: PatentIn version 3.1
SEQ ID NO 87
LENGTH: 2307
TYPE: DNA
ORGANISM: Homo sapiens
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Human Genbank/Y11354.1
DATABASE ENTRY DATE: 1997-06-25
RELEVANT RESIDUES: (1)..(2307)
US-09-893-519A-87

Query Match 5.3%; Score 135.8; DB 9; Length 2307;

Oy	188	TATCTCCTACGAAACGTGTACAGAGCCGAGGCCAAGTAATCACTCAAGGGCAG	247
Oy	51762	TACTTCCAAATTACCAGAACCAAAGAACAACCAAGTAAAATTATTTGGTAGAACAT	5182
Oy	248	CAGTACCAGCGAATCTCTCAACACAGTCGAATAATCTGTACAGTGCGCAACTCTAGTCCACAAAT	307

Db 51822 ATCTTACGAGTATGAGAAATGCAATATATATATTTGGAAATTCACAAACACAC 51881
QY 308 TAATCAAGAAAGTGCAGTGCACCTGTGTAAANA 342
Db 51882 TGATTAGAGAAAAGAACTACTATATATTGANA 51916

RESULT 7
US-09-864-761-10113
Sequence 10113, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-x-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 10113
LENGTH: 486
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL035457.11
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.95
US-09-864-761-10113

Query Match 1.4%; Score 36; DB 10; Length 486;

Best Local Similarity 45.4%; Pred. No. 2.9; Indels 0; Gaps 0;
Matches 129; Conservative 0; Mismatches 155; Indels 0; Gaps 0;
QY 968 AGAAGTGTGTGTCGCAAACTTTGAAACCACTTGTCTGTCAGTGGAGCAAAAGCTGGA 1027
Db 179 AGCTCAGAGCAGCAGCATATCAACACCTGCTCCACGATCTGTATAGCAGCAATACCT 238
QY 1028 GTTGTGACACTTCACTTGTGTGTGGCCCAACTGCTGCAACAGAGGAGCAACAGCTGGA 1087
Db 239 GCTTCAGCACCCTGCATCAGATCAACACCCGCTTCAACATCTGCATCAACATCTGCACCA 298
QY 1088 GGTTCGTCAGACTTCAAAACCACTGTGTGATCTGTGCAAAACAGAGTGCAGCTG 1147
Db 299 GCATCAGTACAGATCAGATGAGTCACTACAGCATCAGATCTGCATCTGCATCAACAT 358
QY 1148 TCACTGCACTGAAAGCCAGTGTCTGTGAAACAGCACTGATCCCTTCAGCA 1207
Db 359 TTTTCAGCAGCTGATCAACACCTGCTTCAACATCTGCATCAACATCTGCACAAAGCTCA 418
QY 1208 GTAACTTTGGAGAACTTCAGTGTGACGCTATTTGTCTTCATC 1251
Db 419 CTACCATCAGCATCCTACACAGCATCATCTGCACCTGCATC 462

RESULT 8
US-09-864-761-18737
Sequence 18737, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: Aeomica-x-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203

;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 18737
;; LENGTH: 510
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL035457.11
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.85
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.6
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.95
;; OTHER INFORMATION: NT HIT: U01287.1, EVALUATE 1.30e+00
US-09-864-761-18737

Query Match 1.4%; Score 36; DB 10; Length 510;
Best Local Similarity 45.4%; Pred. No. 3;
Matches 129; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 968 AGACTGTCTCAGTGCACAACTTTGAACCCACTTGTCTGCTCCAGCTGGGACCAAAAGCTGA 1027
DB 42 AGCTAGAGACCCAGCAGCATCATCAACACCTGCTCAGCATCTGTATCAGCACCAATACCT 101
QY 1028 GTTGTGACACTTCTTGTGGGCCCACTGCTGCACAGAGAAACAGACCTGGAACT 1087
DB 102 GCTTCAGACCTCGCATCAGCATCAACACCCGCTTCAACATCTGCATCAACATCTGCACCA 161
QY 1088 GGTTCCTTCAGACTTCAAAACCACTTGTGACATCTGTGGCAACAGCACTGACACGGTTC 1147
DB 162 GCATCATCTACAGCATCAACATCAGCATCAGCATGAGCATCTGCATCTGCATCAACACAT 221
QY 1148 TCACGCAACCTGAAAGCACTTGTCTGTGGAACAGCATGTAACACTGTCCCTTCAGAGA 1207
DB 222 TTTTCAGACCTCGCATCAACACCTGCTTCAACATCTGCATCAACATCTGCACAAAGCATCA 281
QY 1208 GTAACTTTTGGAAACTTCAGGTGACGCTATTGTCTTCATC 1251
DB 282 CTACCATCAGCATCTACACCATCAGCATGAGCATCTGCACCTGCATC 325

RESULT 9

US-09-864-761-12567
; Sequence 12567, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664

;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 12567
;; LENGTH: 591
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC019141.3
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.68
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.85
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.62
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.69
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.75
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.65
US-09-864-761-12567

Query Match 1.4%; Score 36; DB 10; Length 591;
Best Local Similarity 50.6%; Pred. No. 3.3;
Matches 87; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 1614 AAAATAGATTAAGAGCATTAACATCATCTCCGAGATGAGATGACATCATATGAT 1673
DB 233 AGAAAATAGAGTACGAGAAAACAAACAGCACCAAGAAAACATGACAAATGAAAGTGGAT 292
QY 1674 GTGACTTCTATGCGACACTCATTCAGTCATGTAAGATAAGATGAACCATTTCTTT 1785
DB 293 GTGCTTACAGTACGTGGGTTAACTTAACATATCTTTGTGGGATGGATTAATAATAC 352
QY 1734 TCTGAATTGGTTGGCACACTCATTCAGTCATGTAAGATAAGATGAACCATTTCTTT 1785
DB 353 TCTCAAGTAGATGATACCTGTACCTTCTTACGTGAATACAGATTATTT 404

RESULT 10

US-09-816-894-5/c
; Sequence 5, Application US/09816894
; Patent No. US20020073448A1
; GENERAL INFORMATION:
; APPLICANT: Michalowski, Susan
; APPLICANT: Spiker, Steven
; TITLE OF INVENTION: MATRIX ATTACHMENT REGIONS
; FILE REFERENCE: 5051.401XXDV
; CURRENT APPLICATION NUMBER: US/09/816,894
; PRIOR FILING DATE: 2001-06-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 998
; TYPE: DNA
; ORGANISM: Nicotiana tabacum
US-09-816-894-5

	Query Match	Similarity	1.4%	Score 36	DB 10	Length 998
	Best Local	Similarity	51.2%	Pred. No. 4.6		
Matches	84	Conservative	0	Mismatches	80	Indels 0; Gaps 0
QY	1997	CAAAATTTCTGAAAAAGCTGGATCAATTGTGAGAACGACAGACAAAGATTGTGGAAGAAAGACA	2056			
Db	809	CAAAATTTACTNAAAACTGAAAAAAAACGAAAAATATGTGTTTTCAGATTTTCAAAAAA	750			
QY	2057	AATGTTACTTAAGGACGCCAGACGATCGTTCTAATTAAGCAAGATCCAGAACACTGAGATT	2116			
Db	749	AACCTGCTTAAAGAACTGAAAAATATTTTCTAAATTAATATTTTGTAAAAAACTGAAAAA	690			
QY	2117	AAAGCAGAAAGCCAAAGAGTACACAGATTGGACCTGCACAGA	2160			
Db	689	AAAGCAGAAAGCTAATTTTCTTAAGTAATTAATAAACTGAAAAA	646			

US-09-815-242-3827 RESULT 11
Sequence 3827, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zykkind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes In
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELTRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3827
LENGTH: 2391
TYPE: DNA
ORGANISM: Enterococcus faecalis
US-09-815-242-3827

	Query Match	Similarity	1.4%	Score 36;	DB 10;	Length 2391;
	Best Local	Similarity	51.9%;	Pred.	No. 8.2;	
	Matches	81;	Conservative	0;	Mismatches	75; Indels 0; Gaps 0;
QY	189	ATCTCTCAGCAACTGTGACAGGAGCCGAGACCAAGTAACATNACCCTCAGGCCAGC	248			
Dd	346	AGCAACACACAACAAACCGTAGCAATTCAAAACAATACAAAACCGCATATACCCA	405			
QY	249	AGTACACAGGAATCTCTCAAACAGTCAAAAATCTGTACAGTGCCGAACCTAGTCACAATT	308			
Dd	406	AACATAACGGTTCTACACAGATCAAAATCGCACAGCCAAAATAACAAATGGCGGGAAC	465			
QY	309	AATCAAGAAGTGGCAGTGACACCTCGTTAAAAAATT	344			
Dd	466	AACCAAAACCGTGGCGGCCAAAACCGTAACCACAAT	501			

```

RESULT 12
US-09-815-242-6667
Sequence 6667, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIORITY FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 6667
LENGTH: 2397
TYPE: DNA
ORGANISM: Enterococcus faecalis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(2397)
US-09-815-242-6667

```

```

Query Match          1.4%:  Score 36; DB 10; Length 2397;
Best Local Similarity 51.9%:  Pred. NO. 8.2;
Matches 81; Conservative 0; Mismatches 75; Indels 0; Gaps 0.

OY  189  ATCTCTCAGCAAACTGTAAACAAGACCGAGACCAAGTAACATNACTCTCAAGGCCAGC 248
Db    346  AGCAACACACAACAAACCGTAGCAATCAAAACAATATACAAACAATCAAAACCGCAATACCCA 405

OY  249  AGTACCAGCGAATCTCTCAAAACGTAAATCTGTACAGTGGCCGAACCTCTAGCTCACAATT 308
Db    406  AACATTAACGGGTTCTACAAACGATCAAAATCCGACAGCCAAAATAAACAAATGGCGGGAAC 465

OY  309  AATCAAGAAAGTGGCAGTGCACACTGTTAAAAAATT 344
Db    466  AACCAAAACCGTGGCGGCCAAAACCGTAGACACAAT 501

RESULT 13
US-09-974-300-101
: Sequence 101: Application US/09974300
: Patent No. US20020146721A1
: GENERAL INFORMATION:
:   APPLICANT: Berka, Randy M.
:   APPLICANT: Clausen, Ib Groth
:   TITLE OF INVENTION: Methods For Monitoring Multiple Gene
:   TITLE OF INVENTION: Expression
:   FILE REFERENCE: 10085.500-US
:   CURRENT APPLICATION NUMBER: US/09/974.300
:   CURRENT FILING DATE: 2001-10-05

```

;; PRIOR APPLICATION NUMBER: 09/680,598
;; PRIOR FILING DATE: 2000-10-06
;; PRIOR APPLICATION NUMBER: 60/279,526
;; PRIOR FILING DATE: 2001-03-27
;; NUMBER OF SEQ ID NOS: 8481
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 101
;; LENGTH: 1668
;; TYPE: DNA
;; ORGANISM: Bacillus licheniformis
US-09-974-300-101

Query Match 1.4%; Score 35.8; DB 10; Length 1668;
Best Local Similarity 59.2%; Pred. No. 7.4;
Matches 61; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Oy 1598 AGCATCTCTCTCAGAAAATAGATATAAGAGATGTAATCATGCTTCCGAGTGA 1657
Db 40 AGCAGCTCTCAGTGAAGATATAAGACTTAATAATGTAACGCTTTGCTTTGGCGA 99

Oy 1658 GGATGACATCAATGATGTGACTTCTATGCGAGGGGCTCAACCTT 1700
Db 100 CGGCGCCCAACAACGAGGGGACATTTCAGAGAGGATCAACCTT 142

RESULT 14
US-09-764-878-403/C
;; Sequence 403, Application US/09764878
;; Patent No. US20020090615A1
;; GENERAL INFORMATION:
;; APPLICANT: Rosen et al.
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;; FILE REFERENCE: PA121
;; CURRENT APPLICATION NUMBER: US/09/764,878
;; CURRENT FILING DATE: 2001-01-17
;; Prior application data removed - consult PALM or file wrapper
;; NUMBER OF SEQ ID NOS: 428
;; SOFTWARE: Patentin Ver. 2.0
;; SEQ ID NO 403
;; LENGTH: 2534
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-764-878-403

Query Match 1.4%; Score 35.8; DB 10; Length 2534;
Best Local Similarity 56.3%; Pred. No. 9.7;
Matches 67; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Oy 2001 TTCTTGAAGAGCTGATCAATGTGAGAGCAGAGAAAGATTTGAGAGAAAGAAATG 2060
Db 655 TTTCTTAATAAGAACATTAATAATAAAGCTTAATAAATCTTCTAGTAATAATCAATG 596

Oy 2061 TTACTTAAGCAGCAAGAGCTGTTCTAATAAGAAAGATCCGAGACAGCTGAGATTAA 2119
Db 595 AAAATTATTAATCCAGAAACTTATAGAGAGAAAGTCTTAAGAAATGAATGAGA 537

RESULT 15
US-09-790-988-1
;; Sequence 1, Application US/09790988
;; Patent No. US20020127687A1
;; GENERAL INFORMATION:
;; APPLICANT: SHIGENOBU, SHUJI
;; APPLICANT: WATANABE, HIDEKI
;; APPLICANT: HATTORI, MASAHIRO
;; APPLICANT: SAKAKI, YOSHIYUKI
;; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;; FILE REFERENCE: 081356/0159
;; CURRENT APPLICATION NUMBER: US/09/790,988
;; CURRENT FILING DATE: 2001-02-23
;; PRIOR APPLICATION NUMBER: JP2000-107160
;; PRIOR FILING DATE: 2000-04-07
;; NUMBER OF SEQ ID NOS: 7

;; SOFTWARE: Patentin Ver. 2.1
;; SEQ ID NO 1
;; LENGTH: 640681
;; TYPE: DNA
;; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 1.4%; Score 35.6; DB 10; Length 640681;
Best Local Similarity 46.7%; Pred. No. 3.1e+02;
Matches 113; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

Oy 1769 AGATGACCATTTCTTTTATGAGCTCTACAAAAGAAATCTAGACATTTGTAATA 1828
Db 265588 AAAAGAACTTTAGAAATAAAGTCTTTAAACAGATAGCTTTTGGTATTTTCAAA 265647

Oy 1829 GCATGACATTACAGAACTTAACCTGATGCTGTGAACCTGATCCCAAGCAACAGGA 1888
Db 265648 TGAGTTAATCAAGTAAAAAAATATATCTTTTATTTAGTCTGCCCTGAAAAATTA 265707

Oy 1889 AGCACTAGAGGCTCTAGAAAACTGACTCAATTTGCTCAGATCGAATGACTACTTA 1948
Db 265708 ATTACAAAAGAAATATCAAGTATTTGGAATTTATCTTACAGACATCCATCAATCAATA 265767

Oy 1949 CAGCGAAGTGAATAATTACATCTGTGATGATACAGGTACAGCTCAAAATTTCTTGA 2008
Db 265768 TAAAGAAAGAAATTAAGATGTTATGTAATGTTAAATATATCAGATTAATATCATCA 265827

Oy 2009 AA 2010
Db 265828 AA 265829

Search completed: February 16, 2003, 22:54:38
Job time : 1554 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2003, 22:39:25 ; Search time 3476 Seconds

(without alignments)
11918.313 Million cell updates/sec

Title: US-09-763-909-1

Perfect score: 2558
Sequence: 1 gggaccctgtgaccacaagt.....cttattactcttacctat 2558

Scoring table: IDENTITY_NDC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estda:*
2: em_esthum:*
3: em_estlin:*
4: em_estlnu:*
5: em_estlov:*
6: em_estlpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estlum:*
16: em_estlom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	634.2	24.8	946	14	BQ952832 AGENCOURT
2	537.6	21.0	673	13	BI517527 603041813
3	398	15.6	1116	11	AK012135 Mus muscu
4	395.2	15.4	599	10	BB616334 BB616334
5	370.6	14.5	388	9	AA457205 AA457205
6	344	13.4	427	10	AW297262 AW297262

7	296.4	11.6	570	13	BM539135
8	288.6	11.3	1410	14	BQ220171 AGENCOURT
9	286.4	11.2	787	12	BG213757 RST33376
10	269.4	10.5	2046	11	BC026493 Mus muscu
11	267.8	10.5	777	12	BG217524 RST37234
12	262.4	10.3	507	14	BQ560074 RST37234
13	248.8	9.7	545	10	AM390057 RST37234
14	244	9.5	600	13	BI986567 RST37234
15	241	9.4	734	12	BG259865 RST37234
16	238.8	9.3	940	13	BI152593 RST37234
17	237.6	8.9	404	14	BQ417829 RST37234
18	226.2	8.8	417	9	AL121204 RST37234
19	222.8	8.7	884	9	AL668344 RST37234
20	208.8	8.2	552	10	AM964632 RST37234
21	208.6	8.0	271	12	BG191106 RST37234
22	203.2	7.9	233	12	BG198825 RST37234
23	184	7.2	556	14	BQ560102 RST37234
24	183.6	7.2	518	9	AA438072 RST37234
25	183	7.2	454	12	BE939960 RST37234
26	178.8	7.0	563	12	BE986487 RST37234
27	178.2	7.0	629	14	BM938537 RST37234
28	171.4	6.7	469	12	BE698723 RST37234
29	165.4	6.5	463	9	AA549039 RST37234
30	164	6.4	410	9	AL503691 RST37234
31	161.6	6.3	541	12	BG514288 RST37234
32	161.4	6.3	409	12	BM853763 RST37234
33	158.8	6.2	749	12	BF135081 RST37234
34	158.2	6.2	468	10	AW785378 RST37234
35	154.2	6.0	484	12	BE940003 RST37234
36	151.2	5.9	623	10	AV965961 RST37234
37	148.4	5.8	686	17	AG092765 Pan trogl
38	143.2	5.6	355	10	AM637728 RST37234
39	142	5.6	426	10	BE146234 RST37234
40	141.2	5.5	331	9	AL467327 RST37234
41	135	5.3	919	14	BQ685051 RST37234
42	134.2	5.2	584	10	AW378648 RST37234
43	134	5.2	530	12	BE844520 RST37234
44	130.2	5.1	312	10	AW414768 RST37234
45	128	5.0	282	9	AA360405 RST37234

ALIGNMENTS

RESULT 1
LOCUS BQ952832 946 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCOURT_8949744 NCI_CGAP_Mam2 Mus musculus CDNA clone
IMAGE:6442204 5', mRNA sequence.

ACCESSION BQ952832
VERSION BQ952832.1 GI:22368310
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.
1 (bases 1 to 946)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.jnl.gov
Plate: LLM313966 row: k column: 05
High quality sequence stop: 666.
Location/Qualifiers 1..946

FEATURES
SOURCE

```

/organism="Mus musculus"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:6442204"
/clone_1lb="NCI CGAP Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Salt;
Site_2: Not; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT      322 a      203 c      193 g      226 t      2 others
ORIGIN

Query Match      24.8%; Score 634.2; DB 14; Length 946;
Best Local Similarity 85.0%; Pred. No. 1.9e-169;
Matches 764; Conservative 0; Mismatches 110; Indels 25; Gaps 4;

OY 1409 GTAGTTACAGCCTTCAGAGGCAATGAAAAAGTACCAATTTTCACATTCCTCA 1468
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35 GTGTACACAGCCTTCAGAGGCAATGAAAAAGTACCAATTTTCACATTCCTCA 94

OY 1469 ACATTGACATTCAGAAATGTGAGACAGAAAGACGATGCAAGTAAACATATACCTACT 1528
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 95 CCATTGACACCTCAGAACTGTGGCAGAAAGAC---ACCAATTAATGCTGTGATGCTACT 151

OY 1529 AGTCACTTCTCCAGCTTCATTTCAAGCAAAATTAACCTGCTGGAATTAATTCCT 1588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 152 A-----CTTCATCATTAAGCAAAATTA-CTCTGCTGGAATTAACCTTTT 195

OY 1589 GTCACTTCAAGC-----ATCTCTACTCAGAAAAATAGAAATTAAGAGATTAACATC 1642
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196 GTGCTTCAAGCTCAAAAGATCTTCTTATTCAGACCAATTAATTAAGAGATTAAGACCAAC 255

OY 1643 ATGCTCCGAGATGAGATGACATCAATGATGATCTTATGAGGAGGCTCAACCTTA 1702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 256 ATGCTTCAAGATGAGATGACATCAATGATGATCTTATGAGGAGGCTCAACCTTGA 315

OY 1703 TGAAGAAATGCTGATCTTGAACAACAACTGTAATGTTGGACACTCATTCAGTC 1762
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 TGAAGAAATGCTGATCTTGAACAACAACTGTAATGTTGGACACTCATTCAGTC 375

OY 1763 ATGTAAGATGAACATTTCTTTTATGAGACTCTACAAAGAGAACTTATACATTTG 1822,
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 376 ATGTAAGATGAACATTTCTTTTATGAGACTCTACAAAGAGAACTTATACATTTG 435

OY 1823 TAAAAAGCATGACATTAAGAACTTGAATCTGATGCTGAATGATCTCCCAAGCAAC 1882
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 436 TAAAAAGCATGACATTAAGAACTTGAATCTGATGCTGAATGATCTCCCAAGCAAC 495

OY 1883 ACAGGACACATTAAGAGGCTCTTGAACAACTGACATGCAATGCTCAGCATGAATGAC 1942
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 496 ACAGGACACATTAAGAGGCTCTTGAACAACTGACATGCAATGCTCAGCATGAATGAC 555

OY 1943 TACTTCAAGGGAAGTAAATTAATCTCGTGTAGTATACAGCTCCACAGCTCAAT 2002
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 556 TACTTCAAGGGAAGTAAATTAATCTCGTGTAGTATACAGCTCCACAGCTCAAT 615

OY 2003 TCTTGAAGAGCTGATCAATTTGAGAGCAGAGAAAGATTTGGAAGAAAGATGTT 2062
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 616 TCTTGAAGAGCTGATCAATTTGAGAGCAGAGAAAGATTTGGAAGAAAGATGTT 675

OY 2063 ACTTAAAGCAGCCAGAGTCTTCTAATAAAGAGATCCAGAACAGCTGAGATTAAGCA 2122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 676 ACTTAAAGCAGCCAGAGTCTTCTAATAAAGAGATCCAGAACAGCTGAGATTAAGCA 735

OY 2123 GAAAGCAAGAGTTACAGCAATTTGAGAGATTAAGAGATTAAGAGATTAAGAT 2182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 736 AAAAGCAAGAGTTACAGCAATTTGAGAGATTAAGAGATTAAGAGATTAAGAT 795

OY 2183 CACAGCTCTTGAGATTTGAGCAAGAGAAAGAGACCACTGATTTGGAATTTGAGG 2242
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 796 CACAGCTCTTGAGATTTGAGCAAGAGAAAGAGACCACTGATTAAGATTAAGATGAGAG 855
OY 2243 CTTAAAGACACCTTCTGCTTCTGAGACATCCAGCTGACAGCCACCAAGATTC 2301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 856 CTTTCAAGATTAACCTTCTGAGACTTTCAGGCTTACAGCTTACAGCCCAACATTTTC 914

RESULT 2
BIS17527      673 bp      mRNA      linear      EST 29-AUG-2001
LOCUS      603041813F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5162978 5',
DEFINITION      mRNA sequence.
ACCESSION      BIS17527
VERSION      BIS17527.1 GI:15342319
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 673)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgabbs@remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM1404 row: J column: 03
High quality sequence stop: 673.
Location/Qualifiers
1. 673
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5162978"
/clone_1lb="NIH_MGC_116"
/lab_host="DH10B"
/notes="Organ: pooled colon, kidney, stomach; Vector:
pCMV-SPORT6; Site_1: Not; Site_2: EcorV (destroyed); RNA
source anonymous pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon; 46 yo male kidney, and pool of 2
stomachs, 62 yo male and 70 yo female. Library is
oligo-dr primed and directionally cloned (EcorV site is
destroyed upon cloning). Average insert size 1.4 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
023. Note: this is a NIH_MGC Library."

BASE COUNT      224 a      146 c      137 g      166 t

Query Match      21.0%; Score 537.6; DB 13; Length 673;
Best Local Similarity 98.1%; Pred. No. 6.3e-142;
Matches 565; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

OY 1985 CAGGTACAGCTCAAAATTTCTGAAAGCTGATCAATGGAACAGACGAAAGATTT 2044
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 CAGGTACAGCTCAAAATTTCTGAAAGCTGATCAATGGAACAGACGAAAGATTTG 60

OY 2045 GGAAGAAAGAAATGTTACTTAAGCAGCCAGAGTCTGTTCTAATAAAGATCCAGA 2104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 GAAGCAAGAAAGAAATGTTACTTAAGCAGCCAGAGTCTGTTCTAATAAAGATCCAGA 120

OY 2105 ACAGCTGAGATTAAAGCAGAAAGAGATTACAGCAATTTGGAATTTGACAGATACA 2164
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 ACAGCTGAGATTAAAGCAGAAAGAGATTACAGCAATTTGGAATTTGACAGATACA 180

OY 2165 GCATTAGAGAGCTTAATCTACAGCTCTTGAGATTTGAGCAAGAGAAAGAGACACT 2224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 181 GCATGAGACGCTATCTCACAGCTCTTGACGCTATTGGACCAGGAAGAGACCACT 240

Oy 2225 AGAATCTGGAAATGAGGGCTTAAAGACACCTTCTTCTTGAGGACATCCAGCTGAC 2284

Db 241 AGAATCTGGAAATGAGGGCTTAAAGACACCTTCTTCTTGAGGACATCCAGCTGAC 300

Oy 2285 AGCCACCAAGAGTTGCTGCTCCAGAAATCAGGAATCTGCTCAGGACTTATAT 2344

Db 301 AGCCACCAAGAGTTGCTGCTCCAGGAATCAGGAATCTGCTCAGGACTTATAT 360

Oy 2345 TTGTATGAGACAGAGGAGGATGATATTTCTGAGCTCATACCTGGCCCTCTGAA 2404

Db 361 TTGTATGAGACAGGAGGAGGATGATATTTCTGAGCTCATACCTGGCCCTCTGAA 420

Oy 2405 GTGACCACTCCACTCTTCCATCCACATCCTTGTCTTACTGCCAAGACACAAAGC 2464

Db 421 GTGACCACTCCACTCTTCCATCCACATCCTTGTCTTACTGCCAAGACACAAAGC 480

Oy 2465 ATGTGTCACGTCTCTGAAATTTTCATTTTCGAAAATTA-CACCAACATGAAGAGAT 2523

Db 481 ATGTGTCACGTCTCTGAAATTTTCATTTTCGAAAATTAACCAACATGAAGAGAT 540

Oy 2524 TGTTCACGATAG-AACCTTATTAACCTTACCTAT 2558

Db 541 TGTTCACGATAGAACTTTATTAACCTTACCTAT 576

RESULT 3
AK012135

LOCUS AK012135 1116 bp mRNA linear HTC 19-JAN-2002

DEFINITION Mus musculus 10 days embryo whole body cDNA, RIKEN full-length enriched library, clone:2610524B04:homolog to TRANSCRIPTION INITIATION FACTOR TFIIID 105 KDA SUBUNIT (TAFII-105), full insert sequence.

ACCESSION AK012135 1 GI:12848693

VERSION AK012135.1

KEYWORDS HTC; CAP Trapper.

SOURCE Mus musculus (strain: C57BL/6J) 10 days embryo cDNA to mRNA, clone_11b:RIKEN full-length enriched mouse cDNA library clone:2610524B04.

ORGANISM Mus musculus

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

REFERENCE 2

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20499374

PUBMED 11042159

REFERENCE 3

AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, T., Nishi, K., Kikunaka, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsumoto, Y., Yonekura, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and Hayashizaki, Y.

TITLE RIKEN integrated sequence analysis (RISA) system-384-format

JOURNAL sequencing pipeline with 384 multicapillary sequencer

MEDLINE Genome Res. 10 (11), 1757-1771 (2000)

PUBMED 11076861

REFERENCE 4

AUTHORS Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Atakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S.,

Atakawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamakawa, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nakado, I., Pease, G., Quackenbush, J., Schriml, L.M., Staudt, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, T., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seyer, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyo-Oka, K., Wang, K.H., Weitz, C., Whiteaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohlschl, S., and Hayashizaki, Y.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409 (6821), 685-690 (2001)

MEDLINE 21085660

PUBMED 11217851

REFERENCE 5 (bases 1 to 1116)

AUTHORS Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Atakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bull, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Horii, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.

TITLE Direct Submission

JOURNAL Submitted (10-JUL-2000) Yoshinori Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsukuba, Ibaraki, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

FEATURES

SOURCE

1. 1116

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="PANTOM:DB:2610524B04"

/db_xref="MGD:MGI:1905330"

/db_xref="taxon:10090"

/clone="2610524B04"

/clone_11b="RIKEN full-length enriched mouse cDNA library"

/dev_stage="10 days embryo"

misc_feature

1. 1116

/note="data source:SPRR, source key:Q92750, evidence:ISS homolog to TRANSCRIPTION INITIATION FACTOR TFIIID 105 KDA SUBUNIT (TAFII-105) (TAFII105)"

/db_xref="MGD:MGI:1919754"

BASE COUNT 216 a 347 c 330 g 223 t

ORIGIN

Query Match 15.6%; Score 398; DB 11; Length 1116;

Best Local Similarity 83.5%; Pred. No. 7.7e-102;

Matches 476; Conservative 0; Mismatches 90; Indels 4; Gaps 2;

Oy 1 GGGACCTGGTACCAAGAGTGGCTCGGACAGCCCTCCAAAGTACGAGCGGCTT 60

Db 551 GGGACCTGGTACCAAGAGTGGCTCGGACAGCCCTCCAAAGTACGAGCGGCTT 607

Oy 61 AGGCTGCTGCTCTCAGATAGTCCGCGTGAAGGCCCCCAACACGACGACATCCAGTTT 120

QY	301	CTTGAAATATAACCTCTTACTTCTGCGGACTTCCAGCCTAACAGCCACCAACCAAT	357
QY	2303	TCGTCCAAAGAAATCAGCAAAATCTGCTTCAGGCACTGTGAATATTTGTATGAAACAGAAAG	2362
Db	358	TCGTCCAAAGAAATCAGCAAAATCTGCTTCAGGCACTGTGAATATTTGTATGAAACAGAAAG	417
QY	2363	GGAAGTGAAGATTTCTGCAAGCTCTTACTTGGCCCTTTGTGAAGTACACATCCACTCTTC	2422
Db	418	AGAGATGGAAGATTTCTGCAAGCTCTTACTTGGCCCTTTGTGAAGTACACATCCACTCTTC	473
QY	2423	CATCAACATCTCTGCTTCTTACTGCGCAAGAAAGAACAC--AAAGCATTTGTCACTGTCTT	2480
Db	474	CATCAACATCTCTGCTTCTTACTGCGCAAGAAAGAACACATCTGCGCACTGTCTTCTT	533
QY	2481	GAAATTTCAATTT 2492	
Db	534	GAAACCTCCCTT 545	
RESULT 5			
LOCUS	AA457205		
DEFINITION	AA457205	388 bp	mRNA
VERSION	AA457205		
KEYWORDS	AA457205		
SOURCE	AA457205.1	GI:2179925	
ORGANISM	human.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 388)		
JOURNAL	Hillier, L., Allen, M., Bowles, L., Dubuque, T., Giesel, G., Jost, S.,		
COMMENT	Kueba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,		
	, T., Waterston, R., and Wilson, R.		
	WashU-Merck EST Project 1997		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@watson.wustl.edu		
	This clone is available royalty-free through LLNL; contact the		
	IMAGE Consortium (info@image.llnl.gov) for further information.		
	Seq primer: -28ml3 rev1 ET from Amersham.		
FEATURES			
Source	Location/Qualifiers		
	1..388		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:838408"		
	/clone_lib="Stratagene fetal retina 937202"		
	/sex="mlxed"		
	/lab_host="SOUR (kanamycin resistant)"		
	/note="Vector: pluescript SK-; Site: 1; EcoRI, Sfe, 2;		
	XhoI; Cloned unidirectionally. Primer: 0190 df. Pooled		
	retinal tissue. Average insert size: 1.0 kb; Uni-ZAP XR		
	Vector: -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3'		
	adaptor sequence: 5' CTCGAGATTTTTTTTTTTTTTTT 3"		
BASE COUNT	117 a 88 c 77 g 106 t		
ORIGIN			
Query Match	14.5%	Score 370.6;	DB 9; Length 388;
Best Local Similarity	98.7%;	Pred. No. 2.4e-94;	
Matches 384; Conservative	0; Mismatches 4;	Indels 1;	Gaps 1;
QY	589	CAAGAACTTCTTCTGCAATGTTAATATAAAGTACATGATGGATCAGCAGTCCCTGGAAT	648
Db	1	CAAGAACTTCTTCTGCAATGTTAATATAAAGTACATGATGGATCAGCAGTCCCTGGAAT	60
QY	649	GGGGCAAAATGTGAAGAAAGCTGTGGAAACACTTTTGGATGCAAAATCGAAGCAGAAGA	708

DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	FEATURES	SOURCE
Db	61	GGGTTAAATGTAAGAAAGCTGTGGGAACAATTGGATGCCAAAAATCGAAGCAAGA	120										
Qy	709	ATTTCATGGAAGACGTATGTTGAATCTCAAGCTCTTCAGCTCACCTGTTCCCTT	768										
Db	121	ATTACTACAGGAAGACGTATGTTGAATCTCAAGCTCTTCAGCTCACCTGTTCCCTT	180										
Qy	769	TCTTAAGAAAAGCGTGTGTCCTTACGACAACCTTGCCCTAACCTCCAGAGCTTCATCA	828										
Db	181	TCTTAAGAAAAGCGTGTGTCCTTACGACAACCTTGCCCTAACCTCCAGAGCTTCATCA	240										
Qy	829	GCAATGTTTACGACGACTTCTAGTGAATATGTCATTCCTTACCTGTACTAACAAGTAAAC	888										
Db	241	GCAATGTTTACGACGACTTCTAGTGAATATGTCATTCCTTACCTGTACTAACAAGTAAAC	300										
Qy	889	AACCTTCCTCTGTGCGACAACTACAGCTGCTCCCAAGCAGTCTGAAAAGTCAATTAATGT	948										
Db	301	AACCTTCCTCTGTGCGACAACTACAGCTGCTCCCAAGCAGTCTGAAAAGTCAATTAATGT	360										
Qy	949	TTCTGGAGCAACAGCACCACCAACTGTGT	977										
Db	361	TTCTGGAGCAACAGCACCACCAACTGTGT	388										
RESULT 6	AM297262/c	AM297262	427 bp	mRNA	linear	EST 16-JAN-2000							
LOCUS	UI-H-BM0-a.j1-c-11-0-UI.s1	NCI.CGAP-Sub6	Homo sapiens	cdna	clone								
DEFINITION	IMAGE:2731796	3', mRNA sequence.											
ACCESSION	AM297262												
VERSION	AM297262.1	GI:6703898											
KEYWORDS	EST.												
SOURCE	human.												
ORGANISM	Homo sapiens												
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:												
AUTHORS	Mammalia: Eutheria: Primates: Catarrhini: Hominiidae: Homo.												
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.												
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),												
COMMENT	Unpublished (1997)												
FEATURES	Contact: Robert Strausberg, Ph.D.												
SOURCE	Email: g9apbs-remail.nih.gov												
	The sequence contained an oligo-dT track that was present in the												
	oligonucleotide that was used to prime the synthesis of first												
	strand cDNA and therefore this may represent a bonafide poly A												
	tail. cDNA library preparation: M.B. Soares lab Clone distribution::												
	NCI-CGAP clone distribution information can be found through the												
	I.M.A.G.E. Consortium/LLNL at:												
	www-bio.llnl.gov/dbirp/image/image.html												
	Seq primer: M13 Forward												
	POLYA-Yes.												
	Location/Qualifiers												
	1..427												
	/organism="Homo sapiens"												
	/db_xref="taxon:9606"												
	/clone="IMAGE:2731796"												
	/clone_lid="NCI CGAP-Sub6"			</									

DEFINITION AGNCOURT_7571205 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6052107
5', mRNA sequence.
ACCESSION BQ220171
VERSION BQ220171.1 GI:20401571
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1410)
NIH-MGC <http://mhc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM13307 row: a column: 04
High quality sequence start: 14
High quality sequence stop: 184.
Location/Qualifiers
1..1410
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6052107"
/tissue="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: PCMV-SPORT6; Site:1; NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 346 a 423 c 330 g 307 t 4 others
ORIGIN
Query Match 11.3%; Score 288.6; DB:14; Length 1410;
Best Local Similarity 84.9%; Pred. No. 1.8e-70;
Matches 376; Conservative 0; Mismatches 49; Indels 18; Gaps 4;
QY 1052 CCAACTGCTGCAGAGAGCAACAGTGGTGTCTTCTCAGACTTCAAAACCA 1111
|||||
Db 1 CCAACTGCTGCAC--GGAGGAACACAGCTGGAAGTCTTCTTCAACATTCAAAACCA 58
QY 1112 CTTGTGACATCTGTGGCAACAGTACAGTGTCTCAGTCAACCTGAAAAGCAGTT 1171
|||||
Db 59 CTTGTGACATCTGTGGCAACAGTACAGTGTCTCAGTCAACCTGAAAAGCAGTT 118
QY 1172 GTCCTGGAACAGAGATTAACCTGCTTCCAGAGATTAACCTTGGAGAACTTCAGGT 1231
|||||
Db 119 GTCCTGGAACAGAGATTAACCTGCTTCCAGAGATTAACCTTGGAGAACTTCAGGT 178
QY 1232 GCAGCTATTTCTCTCATCTGTGAAACCTGTGTCTTCTCTGCTGGAGACATCTG- 1290
|||||
Db 179 GCAGCTATTTCTCTCATCTGTGAAACCTGTGTCTTCTCTGCTGGAGACATCTGCA 237
QY 1291 CAAAGCTGTATTTGGAGACTTCAGTTCAAATCAAACTTGGCCAGCGGCGCTGTCTTTC 1350
|||||
Db 238 CAAAGCTGTATTTGGAGACTTCAGTTCAAATCAAACTTGGCCAGCGGCGCTGTCTTTC 297
QY 1351 ACAACAGCTGGAGATTCACAGGAGAGTTCAGGAGCAAGCACTATTTCTATTGTTTCACT 1410
|||||
Db 298 ACAACAGCTGGAGATTCACAGGAGAGTTCAGGAGCAAGCACTATTTCTATTGTTTCACT 343
QY 1411 AGTTCAGAGCTTCAGAGGAGCAATGAAGAAAGAGTACCAATTTTCAATTCCTCAAC 1470
|||||
Db 344 AGTTCAGAGCTTCCTGTGAGCAGCAGTGAAGAAAGCAAGCAGCAGCTTTTCAATTTCCCAAC 403

QY 1471 ATTGACATTCAGAAATGTGCAC 1493
|||||
Db 404 ATTGACATTCAGAAATGTGCAC 426
RESULT 9
LOCUS BG213757 787 bp mRNA linear EST 21-APR-2001
DEFINITION R5733376 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG213757
VERSION BG213757.1 GI:13735444
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 787)
Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramchandran,R., Whittington,J.,
Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith,
E., Veloso,N., Kilka,A., Hess,J., Cothren,K., Lo,K., Offenbacher,
J., Dandig,J. and Ducar,M.
J., Dandig,J. and Ducar,M.
TITLE Creation of genome-wide protein expression libraries using random
activation of gene expression
JOURNAL Nat. Biotechnol. 19 (5), 440-445 (2001)
MEDLINE 21227151
COMMENT Contact: Scott J. Cain
Athensys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scaine@athensys.com
High quality sequence stop: 443.
Location/Qualifiers
1..787
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Athensys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression', the
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
BASE COUNT 223 a 176 c 160 g 227 t 1 others
ORIGIN
Query Match 11.2%; Score 286.4; DB:12; Length 787;
Best Local Similarity 97.5%; Pred. No. 4.9e-70;
Matches 312; Conservative 0; Mismatches 6; Indels 2; Gaps 2;
QY 2241 GGCTTAAGCAACCTCTTCTCTGAGCATCTCAGCCTGACAGCCACCAACAGTTG 2300
|||||
Db 4 GGGCTTAAGCAACCTCTTCTCTGAGCATCTCAGCCTGACAGCCACCAACAGTTG 63
QY 2301 CATGCTCAAAATACAGAGATCTGCTCAGAGGACTGATATTTTGTATGGAACAGGAA 2360
|||||
Db 64 CATGCTCAAAATACAGAGATCTGCTCAGAGGACTGATATTTTGTATGGAACAGGAA 123
QY 2361 CGGAGATGAAGTATCTCGAGCTCTATACCTGGCCCTTCTGAAGTGACCACTCACTCT 2420
|||||
Db 124 CGGAGATGAAGTATCTCGAGCTCTATACCTGGCCCTTCTGAAGTGACCACTCACTCT 183
QY 2421 TCATCCAGATCTCTGCTATTTACTGCGCAAGAGACCAAGCAATTTTGTGACTGTCT 2480
|||||
Db 184 TCATCCAGATCTCTGCTATTTACTGCGCAAGAGACCAAGCAATTTTGTGACTGTCT 243
QY 2481 GAAATTTCAATTTCTGAAATTA--CACCAATGAAGAGCAATTTTGTAGATTAG--AA 2538
|||||
Db 244 GAAATTTCAATTTCTGAAATTAATCAACCAATGAAGAGCAATTTTGTAGATTAGAA 303
QY 2539 CTTTATTAACCTTACCTAT 2558

|||||

Db 304 CTTATTACTTACTTACTTAT 323

RESULT 10
BC026493

LOCUS
DEFINITION
BC026493 2046 bp mRNA linear HTC 07-AUG-2002
Mus musculus, similar to TAR4 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 135 kD, clone IMAGE:4505179, mRNA.

ACCESSION
BC026493

KEYWORDS
GI:20072943

SOURCE
house mouse.

ORGANISM
Mus musculus

REFERENCE
AUTHORS
TITLE
JOURNAL
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2046)
Direct Submission
Submitted (02-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: The Cepro Laboratory
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) medepax11.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

REMARK
COMMENT
This clone has the following problem: incomplete processing.

FEATURES
source
1..2046
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:4505179"
/issue_type="Eye, retina, mouse straln C57BL/6"
/lab_host="NIH MGC_94"
/note="Vector: pCMV-Sport6"
BASE COUNT 609 a 429 c 466 g 542 t
ORIGIN

Query Match 10.5%; Score 269.4; DB 11; Length 2046;
Best Local Similarity 63.2%; Pred. No. 7.3e-65;
Matches 439; Conservative 0; Mismatches 241; Indels 15; Gaps 1;

QY 1730 AACTGTGAATGGTGGCACTCATTCAGTCATGAAGATGAACCATTTCTTTTAT 1789
|||||

Db 22 AACTGTGAATGGTGGGAGGAGTGAACCCGATCTGTAAAGAGACACCTTCCGCC 81
|||||

QY 1790 TGAAGCTCTACAAAGAGATCTTACACATTTGTAAGAACATGACATTAACACTTAA 1849
|||||

Db 82 CGGCGCTTTGCAAGAGGATCTAGTGATGATTAAGAACAGCAGCATCAGGAACTGCA 141
|||||

QY 1850 CTGTGAGCTGTAAGTGTATCCCAAGCAACAGAGAGAGTACAGAGCTTGTAGA 1909
|||||

Db 142 TCCAGATGTAGTACGTCTGTCTATGCGACACAGCAAGGCTCGAGATCTTGTAGA 201
|||||

QY 1910 AAAACTGACTGCAATTTGCTGACGATGATGACTTACAAAGCAAGTGAATATTCAT 1969
|||||

Db 202 AAAAATCTGMAAGCGCTACGAGAAAACTTCTTTATTAAGACGACGATACGA 261

QY 1970 CTTGTGTAGTATACAGAGTACAGCTCAATTTCTTGAAGAGTGATCAATTTGAGAA 2029
|||||

Db 262 GCAAGCGATGATGTCGGGCTCAGCTCAATTTCTTGAACCACTGGATCAGTTGAAA 321
|||||

QY 2030 GCAGAGAAAGATTTGGAGAAAGAAATGTTACTTAAGCGACCCAGAGCTGTTAA 2089
|||||

Db 322 GCAGAGAAAGATGAGCAGAGAGGAGATTTAATCGGCGACGCAAGTCTGATTCAG 381
|||||

QY 2090 TAAAGAGATTCAGAAACAGCTGAGATTAAGCAAGAAACCAAGAGTTACAGCAATTCGA 2149
|||||

Db 382 ACAAGAAAGACCCAGAGCAATTAAGTTGAGCAAGAAAGAGATGACAGCAGCA 441
|||||

QY 2150 ACTTGCACAGATACAGCATAGAGACGCTAATCTCACAGCTTTGACGCTTTAGACCAAG 2209
|||||

Db 442 GCTGGCAGACATGGGACCAAGAGATGTCACACCTCAGCTGCGAGCGATCGGGCCAG 501
|||||

QY 2210 GAAGAGAGACCA-----CTAGATCTGGAATTTGAGGGCTTTAAAGACAA 2254
|||||

Db 502 GAAAGAGAGAGAAAGTGACATGCACTGGCCAGCATGAGAGAGAGCGCTCAGCGCCG 561
|||||

QY 2255 CTTTCTTGTCTGGGACATCCAGCCTGACAGCCCAACAGATGATGCTCCAGAAAT 2314
|||||

Db 562 CGCAGCAGTCCCGAGCGGCTCCGGCTGGAGAACCCCGACAGTTTCACACGGCAAGAAAT 621
|||||

QY 2315 CACGAGATCTGCTCAGGAGCTTGATATTTTGTATGGAACAGAAACGCGAGATGAATGA 2374
|||||

Db 622 CACGCGGCTCAACTCAGGACGCTCATATTTTGTGAAGAACAGCGTGACAGCAAGCCA 681
|||||

QY 2375 TTCTCGAGCTCTATCTGCGCTTCTGAAAGTGAC 2409
|||||

Db 682 TTCACGTGTGCTCTACAAAGCATTCCTTAAGTAAC 716
|||||

RESULT 11
BG217524

LOCUS
DEFINITION
BG217524 777 bp mRNA linear EST 21-APR-2001
RST37234 Athersys RAGE library Homo sapiens cdna, mRNA sequence.

ACCESSION
BG217524

VERSION
BG217524.1 GI:13743545

KEYWORDS
EST.

SOURCE
human.

ORGANISM
Homo sapiens

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
Cain, S., Leventhal, C., Thornton, M., Ramchandran, R., Whittington, J.,
Lerner, L., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith,
J., Danzig, J., Hess, J., Cochran, K., Lo, K., Offenbacher,
J., and Ducar, M.
Creation of genome-wide protein expression libraries using random
activation of gene expression
Nat. Biotechnol. 19 (5), 440-445 (2001)
21227151

JOURNAL
MEDLINE
COMMENT
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scaineatersys.com
High quality sequence stop: 549.
Location/Qualifiers
1..777
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-Wide Protein Expression Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation

LOCUS	734 bp	mRNA	linear	EST 21-FEB-2001
DEFINITION	6023933377F1 NIH_MGC_94 Mus musculus CDNA clone IMAGE:4505179 5', mRNA sequence.			
ACCESSION	BC295865			
VERSION	BC295865.1 GI:13057927			

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
1 (bases 1 to 734)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM10378 Row: 1 Column: 20
High quality sequence stop: 721.
Location/Qualifiers
1. 734
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:4505179"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 223 a 174 c 211 g 126 t
ORIGIN

Query Match 9.4%; Score 241; DB 12; Length 734;
Best Local Similarity 62.4%; Pred. No. 4.4e-57;
Matches 434; Conservative 0; Mismatches 245; Indels 16; Gaps 3;

QY 1730 AAACCTGAAATTGGTGGCACACATCATTCATGTAAGATGAGACCATTCCTTTAT 1789
DB 21 AAACCTGAGTGTGGGGACGCTGACCCGATCCGTAAAGACACACCTTCCTCC 80
QY 1790 TGGAG-CTCTACAAAAGAACTTAGACATTTGTAAGAACATGACATTCAGAACTTA 1848
DB 81 CGTGGCCTTCGAGAGAGAGATGAGATAGTAAGAAAAGCAGCATCAGGAACTGC 140
QY 1849 ACTCTGATGCTGTGAACTTGATTCGCCAAGCAACAGAGAAAGACTAGAGGCTTTAG 1908
DB 141 ATCCAGATGTAGTGTACGTCTCTATGCCACACACAAAGGCTCAGAAATCTGTAG 200
QY 1909 AAAAAGTGCATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1968
DB 201 AAAAATATCTGAAACGGCTGACGAAAAATCTCTTATAGGACGACGACGATACG 260
QY 1969 TCCTGTAGTATACAGTACAGTCAATTTCTTGAAGAGTGTATGATGATGATGATGAT 2028
DB 261 AGCAAGGAGTATGTCGGGCTCAGCTCAATTTCTTTGAACAACTGATGATGAAA 320
QY 2029 AGCAGAGAAAGATTTGAGAGAAAGAAATGTTACTTAAGCAGCCAAAGAGTCTCTA 2088
DB 321 AGCAGAGAAAGATGAGAGAGAGAGATCTTAATCGGGCAGCGAAGTCTGATCCA 380
QY 2089 ATAAAGAAATCCAGAACAGTGAATTAAGCAAAAGCCAAAGAGTTACAGCAATGG 2148
DB 381 GACAGAGAACCCAGAGCAATTAAGGTGAAGCAAAAGAGATGACAGCAGAG 440
QY 2149 AACTTGCACAGATACAGATAGAGAGCTAATCTCACAGCTCTTGACCTATTGGACCA 2208
DB 441 AGCTGGCACAGATGAGCAAGAGATGCCAACCTCACATGCACTGACAGAGATCGGCCCA 500
QY 2209 GGAAGAAAGACCACTAGAAATCTGAAATTTGAGGCTTAAAGACAACTCTTCTTCG 2268

DB 501 GGAAGAAAGAGAAAGTGTGATGACATGCGGCCAGGATCAGAGACAGAGGGCTCAGGCCCG 560
QY 2269 GGACA-----TCAGGCTGACAGCCACCAAGAGTTGATTCGTCGAAGAT 2314
DB 561 GCGCAGCAGTCCACAGCGCGGCTCGGGAACCCCGACAGATGTCACACGCGCAAGAT 620
QY 2315 CACGAGAAATCTGCTCAGGAGCTTGATTTTGTATGGAACAGAAACGAGATGAAGTA 2374
DB 621 CACGGGGTCAACCTCAGAGGAGCTGATA-TTGTGTAGAAAAGCAGCGTGAGACAGCA 679
QY 2375 TTCTGAGCTTATACCTGCGCCCTTCTGAAGTAC 2409
DB 680 TTCACCTGTGCTTACAAAGCATTCCTTAAGTAAC 714

Search completed: February 17, 2003, 02:56:40
Job time : 3519 secs

This Page Blank (uspto)

XX Dikstein R, Yamit-hezi A;
 XX WPI; 2000-256640/22.
 DR N-PSDB; AA290465.
 XX
 PT Polypeptide encoding TATA box binding protein associated factor II 105
 PT useful for treating e.g. cancers and inducing apoptosis has a dominant
 PT negative effect on the normal biological activity of the binding
 PS protein -
 PS
 PS Claim 7: Fig 2: 48pp: English.
 CC This represents a polypeptide comprising a (modified) fragment (I) of
 CC a TATA box-binding protein associated factor II 105 (TAFII105). A
 CC pharmaceutical composition comprising (I) or the polynucleotide or an
 CC inhibitor or antagonist of (I) is useful for treating cancers and
 CC inducing apoptosis in pathological cells. The composition is also useful
 CC for treating autoimmune diseases, inflammatory processes and viral or
 CC bacterial infections.
 CC
 XX Sequence 852 AA;
 SQ
 Query Match 99.9%; Score 4258; DB 21; Length 852;
 Best Local Similarity 99.9%; Pred. No. 9.2e-278;
 Matches 851; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 661 TRSOLKLEKLDLEKORRLEEREMILKAASRSNKEDEPOLRLKOKAKELQOLELAQI 720
 QY 721 QHRDANLTATAIGPRKKRPLESGIEGLKNDNLASGTSSTLTATKOLARRPRTICRLDLI 780
 DB 721 QHRDANLTATAIGPRKKRPLESGIEGLKNDNLASGTSSTLTATKOLARRPRTICRLDLI 780
 QY 781 FCMQEREMKYSRNLALALKZPLHSSIHILAIYCORRHALLHCEPISISGKZHOHERA 840
 DB 781 FCMQEREMKYSRNLALALKZPLHSSIHILAIYCORRHALLHCEPISISGKZHOHERA 840
 QY 841 LFTIRTLTLTLTY 852
 DB 841 LFTIRTLTLTLTY 852
 RESULT 2
 AAM31494
 ID AAM31494 standard; protein; 801 AA.
 XX
 AC AAM31494;
 XX
 DT 28-APR-1998 (first entry)
 XX
 DE Human hTAFII105 protein.
 XX
 DE TATA-binding protein associated binding factor 105; human; activator;
 KW hTAFII105; transcription factor; TFIID; transcriptional activation;
 KW antibodies; diagnosis; therapy; biopharmaceutical industry.
 XX
 OS Homo sapiens.
 XX
 PN US5710025-A.
 XX
 PD 20-JAN-1998.
 XX
 PF 02-OCT-1996; 96US-0725012.
 XX
 PR 02-OCT-1996; 96US-0725012.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Dikstein R, Tjian R;
 XX
 DR WPI; 1998-109818/10.
 DR N-PSDB; AA002872.
 XX
 PT DNA encoding human tata-binding protein associated factor - for
 PT producing recombinant protein
 PS
 PS Claim 1; Col 17-22; 12pp: English.
 XX
 CC This cDNA sequence represents a human tata-binding protein associated
 CC factor, hTAFII105, isolated from Daudi cell nuclear extracts. Tightly
 CC associated subunits (TAF's) are components of the transcription factor
 CC TFIID and are thought to mediate transcriptional activation. This encoded
 CC protein may be produced recombinantly from transformed host cells or
 CC purified from human cells. hTAFII105 specific binding agents such as
 CC specific antibodies could be used for diagnosis (e.g. genetic
 CC hybridisation screens for hTAFII105 transcripts), therapy (e.g. gene
 CC therapy to modulate hTAFII105 gene expression) and in the
 CC biopharmaceutical industry (e.g. as immunogens, reagents for isolating B
 CC cell specific activators or other transcriptional regulators).
 XX
 SQ Sequence 801 AA;
 Query Match 93.6%; Score 3990; DB 19; Length 801;
 Best Local Similarity 99.9%; Pred. No. 9.1e-260;
 Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;


```

OY 421 VSFCDHICKPVIGTPVQIKLAQPCPVLSQAPGIPITGSSSKOLFSLFHVQOPSGGENK 480
DB 421 VSFCDHICKPVIGTPVQIKLAQPCPVLSQAPGIPITGSSSKOLFSLFHVQOPSGGENK 480
OY 481 QVTTSHSSSTLTIOKCGOKMPVNTIIPISQPPASTIKOITLPGNKITLSLOASPTQKNR 540
DB 481 QVTTSHSSSTLTIOKCGOKMPVNTIIPISQPPASTIKOITLPGNKITLSLOASPTQKNR 540
OY 541 IKENVTSCFEDEDINDVYTSMAGVNLNEENACILATNSELVGTLIOSCKDEPEFLGALQ 600
DB 541 IKENVTSCFEDEDINDVYTSMAGVNLNEENACILATNSELVGTLIOSCKDEPEFLGALQ 600
OY 601 KRLLDGGKHDITELNSDAVNLISQATQERLCKLEKTLIAOHRMTTYVASENYIICSD 660
DB 601 KRLLDGGKHDITELNSDAVNLISQATQERLCKLEKTLIAOHRMTTYVASENYIICSD 660
OY 661 TRSOLKFLKLDLEKORNDLEERMLKAKRSNKDEDEOLRLKOKAELOOLEAQI 720
DB 661 TRSOLKFLKLDLEKORNDLEERMLKAKRSNKDEDEOLRLKOKAELOOLEAQI 720
OY 721 QHRDANLTATAIGPRKKRPLESGIEGLKNDNLASGTSSTLATKOLHRPRITRICLDLI 780
DB 721 QHRDANLTATAIGPRKKRPLESGIEGLKNDNLASGTSSTLATKOLHRPRITRICLDLI 780
OY 781 FCMEOEREMKYSRATYLLALK 801
DB 781 FCMEOEREMKYSRATYLLALK 801

RESULT 4
AAR56494
ID AAR56494 standard; protein; 737 AA.
XX
AC AAR56494;
XX
DT 23-MAR-1995 (first entry)
XX
DE TATA-binding protein-associated factor hTAFl1130.
XX
KW TATA-binding protein associated factor; hTAFl1130; screening;
XX diagnostic; therapeutic; gene transcription regulation.
XX
OS Homo sapiens.
XX
PN WO9417087-A.
XX
PD 04-AUG-1994.
XX
PF 28-JAN-1994; 94WO-US01114.
XX
PR 28-JAN-1993; 93US-0013412.
XX
PR 30-JUN-1993; 93US-0087119.
XX
PA (REGC ) UNIV CALIFORNIA.
XX
PI Comel L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
PI Tjian R, Wang E, Weinzierl KOJ;
XX
DR MPI; 1994-264019/32.
XX
DR N-PSDB; AAQ70731.
XX
PT TATA-binding protein associated protein factors - and
PT corresponding nucleotide sequence and deriv. antibodies, useful
PT in screening, diagnostics and therapeutics
XX
PS Disclosure; Page 142; 180pp; English.
XX
CC The TATA-binding protein associated factor hTAFl1130 (including
CC specific antibodies and fusion products) are used in drug screening,
CC diagnostics and therapeutics. They are used in the development of
CC specific biochemical assays for screening compounds that agonise or
CC antagonise selected transcription factors involved in regulating

```

```

CC gene expression associated with human pathology.
XX
SQ Sequence 737 AA;
Query Match 30.7%; Score 1307.5; DB 15; Length 737;
Best Local Similarity 40.0%; Pred. No. 2,4e-79;
Matches 330; Conservative 103; Mismatches 200; Indels 193; Gaps 22;
OY 13 PPRVSSG-----PRLPAPQIVAVKAPMTTITQFPANQLPPTGLIKNSGFLM 61
DB 68 PTAFTSGIRATLTPVLAAPLPQ-----PQNPRTNQ-----NQLPQGMVLVSENGQL 118
OY 62 LVSPQQTNR-----AETTSNITSRPAPNDQVAKICTVPNSSQLIKRVATPKYL 115
DB 119 MI-POCALMOQAHOAHQOPOTMAPRAPPTSAAPVOISTVQAPGPPIIAR-GVTP---- 172
OY 116 AQGTGVVTVTPKPSVQSAVPTSVTVTPKRPINVT--TLKPSLGS--TPS 168
DB 173 ----TTIKOV---SOAQTVQPSATLQSPGVQPOLVIGGAQVMSLGTATAVGTGPQ 225
OY 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFAMILIKLACSGSGSPSEMGOAKTKIVEOL 227
DB 226 RTVPGATTTSSATE-----TMENVKCKNFISTLIKLASGSKSTETANKEVLONL 279
OY 228 LPAKTEAEETRKLYELKSSPPHLYPFLKKSVALROLPLPSOSFTQOCVQ---TS 283
DB 280 LDKETAEDEFTSRLYRELNSPPQYLVPLFKRSILPALROLTPDSAFIQSOQOQPPPT 339
OY 284 SDVVIATCTTGTTPVTTVTSVSSQSEKSIISGATPAPTVSVQTLNPLAGPVAKAGV 343
DB 340 Q-----ATTALTAVLVSSSVQRTACKTAATVTSALQPPVLSL----- 376
OY 344 VTLHVGPRATGAGTLAGTLLQTSKPLVTSVANTVTVSLQPEKPVSGTAVTSLPAY 403
DB 377 -----TQPGVGVKQGOPTPLVIG-----QPPKP-----GALIRPQY 410
OY 404 TFGETSGAATCLPSVKPVVSCFDHICKPVIGTPVQIKLAQPCPVLSQAPGIPITGSSSKO 463
DB 411 TLTQT-----PWVALRQPH-NRIMLTTPQOQL----- 437
OY 464 LFSLFHVQOPSGGENKQVTTISHSSTLTIOKCGOKMPVNTIIPISQPPASTIKOITL 523
DB 438 -----NPQAPVVVAVPANY 451
OY 524 PGNKIL---SLQASPTQKNRIKENVTSCFEDEDINDVYTSMAGVNLNEENACILATNSEL 580
DB 452 PGTKALSAVSAQAAMAKNKKKEPGGSGFRDDINDVASMAGVNLSEESARLATNSEL 511
OY 581 VGTLIOSCKDEPEFLGALQKRLDGGKHDITELNSDAVNLISQATQERLCKLEKTLIA 640
DB 512 VGTLTRSCKDETEFLQAPLQRLILEIGKKHGTTELHPDVVSYSVSHATQORLQNLVERKISE 571
OY 641 IAOHRMTTYVASENYIICSDPTRSOLKFLKLDLEKORNDLEERMLKAKRSNKDEP 700
DB 572 TAOQKNFSYKDDRDYEDASVYRQQLKFEQLDIEKORNDLEERMLKAKRSNKDEP 631
OY 701 EQLRLKOKAKELQOLELAQIOHRDANLTATAIGPRKKRPLE-----SGIEGLKNDLNL 755
DB 632 EQLRLKOKAKEMQOQELAQORORDANLTATAIGPRKKRKYDCPGSGSGEGSGESVYP 621
OY 756 GTSSTLATKOLHRPRITRICLDLIFCMDEEREMKYSRATYLLALK 801
DB 692 GSSGVGTPRQTRQTRITRVNLRLIFCLENERETSHSLLYKAFK 737

RESULT 5
AAM06084
ID AAM06084 standard; protein; 737 AA.
XX
AC AAM06084;
XX
DT 27-JAN-1997 (first entry)
XX

```

DE Human TATA-binding protein associated factor hTAFII130 protein.
 XX
 XX Human; TATA-binding protein; TBP associated factor; TFIID; holoenzyme;
 KW RNA polymerase II; transcription; messenger RNA; nuclear fraction;
 KW lambda-gt11; expression library.
 XX
 OS Homo sapiens.
 XX
 PN US534410-A.
 XX
 PD 09-JUL-1996.
 XX
 PF 28-JAN-1993; 93US-0013412.
 XX
 PR 28-JAN-1994; 94US-018582.
 PR 28-JAN-1993; 93US-0013412.
 PR 30-JUN-1993; 93US-0087119.
 XX
 PA (REGC) UNIV CALIFORNIA.
 PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
 PI Tjian R, Wang E, Weinzierl ROJ;
 XX
 DR WPI; 1996-333245/33.
 DR N-PSDB; AAT42217.
 XX
 PT Screen for cpts. that bind human TATA-binding protein associated
 PT factor - by testing ability to bind to polypeptide fragments of the
 PT factor, useful as (ant)agonists of transcription factors involved in
 PT disease.
 XX
 PS Examples; Column 105-112; 86pp; English.
 XX
 CC This is the amino acid sequence of the human TATA-binding protein (TBP)
 CC associated factor (TAF) designated TAFII130. The protein is a component
 CC of the TFIID fraction required for reconstituting RNA polymerase II in
 CC vitro transcription activity. The encoded protein has an estimated mol.
 CC wt. of 130 kD by SDS-PAGE.
 CC The invention relates to purified proteins involved in transcription
 CC by RNA polymerase II, the RNA polymerase which transcribes messenger
 CC RNA. RNA polymerase II transcription proceeds in vitro upon addition
 CC of several nuclear fractions designated TFIIA, B, D, E, F, H, I and J
 CC to RNA polymerase II holoenzyme. Fraction TFIID has been shown to
 CC contain a TBP and other TAFs. Purification of TFIID and separation of
 CC its components reveals 7 proteins ranging in size from 30-250 kD.
 CC Serum raised against the TFIID fraction allowed cloning of the corresp.
 CC genes from lambda-gt11 expression libraries.
 CC
 XX
 SQ Sequence 737 AA;
 Query Match 30.7%; Score 1307.5; DB 17; Length 737;
 Best Local Similarity 40.0%; Pred. No. 2.4e-79;
 Matches 330; Conservative 103; Mismatches 200; Indels 193; Gaps 22;
 QY 13 PPKVSSG-----PRLPAQIVAVKAPNTTIOFPANLQUPGTVLIKSNGLM 61
 DB 68 PTATSGIRATLPTVLAPRLQP-----PQNPTNIO---NFGPPGWLVRSGNQL 118
 QY 62 LVSPOQYTR-----AETSNITSRPAVPANPQVTKICTYVNSSQLIKKAVTPVKL 115
 DB 119 MI-PQOALAOQAOAHAPQTTMAPRPATPTSAPVQISVQAPPTPIAR-QVNP---- 172
 QY 116 AIGTIVTVTPKPSVSVAVPTSVTPCKPLNTVT--TLKSSISGASS-----TPS 168
 DB 173 ----TTIIKQV---SQAGTTPQPSATLORSFVQVLVIGGAOYASLGTATAVGTGTPQ 225
 QY 169 NE-PULKAENSAVQINISPTMLENVKCKKFLAMLIKLAGSGSPKMGVKKLVQQL 227
 DB 226 RTVPQATTTSSAATF-----TMENVKCKKFLSTLIKLAGSGKSTETAAVAKELVQNL 279
 QY 228 LDARTAEAEFRKLVELKSSPQHLVPLKKSVALQQLPNSQSPFIQQCVQO-----TS 283
 DB 280 LDGKTEADEFTSRILXRELINSSPQPLVPLFKLSLALROLPPDSAAFTIOOQOQPPPPPS 339

QY 284 SDMVATCTTIVTTPSPVTVTVSSQSEKSIISGATAPRTVSGTFLNPLAGPVGAKAGV 343
 DB 340 Q-----ATTALTAIVLVSSVQRTAGTAATVATLALQPVLSL----- 376
 QY 344 VTLHSGPTAANGTGTAGTGLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAVTSLP 403
 DB 377 -----TQPTQVGKQGGQPTPLVIO-----QPKP-----GALIRPQV 410
 QY 404 TFGETSGAALICLPKPVVSCWDHICKPVIGTPIQIKLADGVPVLSQPAIGTSSSKQ 463
 DB 411 TLTOR-----PVALRQPH-NRIMLTTPQOIDL----- 437
 QY 464 LFSLFHVQVQPSGNEKQVTTISHSSSTLQCGGKTPVNTIPTSQFPASILKQITL 523
 DB 438 -----NPLQPVVKKPAVL 451
 QY 524 PGNKIL---SLOASPTQNRKIKENYTSQFREDINDYTSMAVNLNEENACILATNSEL 580
 DB 452 PCKTALSAVSAQAAAKKKEPEGGSGFRDDDDINDVASMGVNLSESAIILATNSEL 511
 QY 581 VGTLLQSKDEFFLTGALQKRIIDIGKKHDTTELSNDAVNLISQATQERLGLLEKLT 640
 DB 512 VGTLLRSCKDEFTLLQAPLQRRILEIGKHGTTLELHPVSVYSHATQRLQNLVEKISE 571
 QY 641 IAOHMTYKASENYILCSDFRSQKLFLEKLDLEKORRDEERMLKAASRSNKEDP 700
 DB 572 TAOQNFYSKDDRIEASDVAAQKFFEQLDQIKOKDEKERILMAKSRQEDP 631
 QY 701 EQLRLKQAKELQOLELAQIOHRDANLTATAIGPKRRKPLE----SGIEGLKDNLLAS 755
 DB 632 EQLRLKQAKELQOLELAQIOHRDANLTATAIGPKRRKRYDCCPPGSGAEGSGSV 691
 QY 756 GISSLITAKQLHRPRITRCLDLIFCMEQREEMYSALYIALLK 801
 DB 692 GSSGVTGPTQFTRORITRYNLDLIFCENERTSHSLLYKAFK 737
 RESULT 6
 ID AAM25019
 AAM25019 standard; Protein: 737 AA.
 AC AAM25019;
 XX 08-OCT-1997 (first entry)
 DE TATA-binding protein associated factor, hTAFII130.
 XX
 KW TATA-binding protein associated factor; TAF; nuclear protein;
 KW RNA polymerase transcription; TATA-binding protein; TBP;
 KW initiation.
 OS Homo sapiens.
 XX
 PN US5637686-A.
 XX
 PD 10-JUN-1997.
 XX
 PF 28-JAN-1993; 93US-0013412.
 XX
 PR 28-JAN-1994; 94US-018582.
 PR 28-JAN-1993; 93US-0013412.
 PR 30-JUN-1993; 93US-0087119.
 PR 09-MAY-1996; 96US-0646715.
 XX
 PA (REGC) UNIV CALIFORNIA.
 PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
 PI Tjian R, Wang E, Weinzierl ROJ;
 XX
 DR WPI; 1997-319113/29.
 DR N-PSDB; AAT79595.
 XX

PT Nucleic acids encoding human TATA-binding protein associated factor
 PT (TAF) peptide(s) - for production of recombinant peptide(s), used
 PT for modulating transcription of TAFs
 XX Claim 1; Column 111-116; 86pp; English.
 CC AAW52018 represents TATA-binding protein associated factor (TAF)
 CC polypeptide, hTAF1100 (mol. weight 100kd). TAF peptides derived
 CC from hTAF1100 alpha, hTAF1130 beta, hTAF1140, hTAF1170, hTAF1100,
 CC hTAF1130, hTAF1250, hTAF1148 and hTAF110 and nucleic acids encoding
 CC them, are used to modulate transcription, including transcription
 CC initiation. TAFs are nuclear proteins involved in RNA polymerase I,
 CC II and III transcription. The peptides act by binding to a different
 CC TAF, an activator, or TBP (TATA-binding protein) or competitively
 CC inhibiting association of a TAF domain with another compound, typically
 CC a protein like TBP or another TAF, an activator, or DNA.
 CC
 XX
 SQ Sequence 737 AA;
 Query Match 30 7%; Score 1307.5; DB 18; Length 737;
 Best Local Similarity 40.0%; Pred. No. 2,4e-79;
 Matches 330; Conservative 103; Mismatches 200; Indels 193; Gaps 22;
 13 PPKVSSG-----PRLPAPQIVAVKAPVTTTQFPANLQLPQGVILKISNGSLM 61
 68 PRTATSGIRATLPVLAAPRLPQ-----PQNPNTQ---NQLPFGMTLVSENGQL 118
 62 LVSPOQTWR-----AETTSNITSRPVAPNPQVICTVPSNSQLIKVAVTPVKL 115
 119 MI-POALAQMOQAHAQOPQTMAPRPAPTSAPVQISTVQAPGRPIAR-QVTP---- 172
 116 AQGTIVTVTPKPSVQSAVPTSVTVTPGKPLNTVT--TLKPSLSGSS-----TPS 168
 173 ----TTLIKOV---SQAQTVVQPSATLQSRSPGVQPOLVIGCAQATLGTATAVQTGPQ 225
 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFKLMLIKLACSGSOSQSEMOQNKYLEOL 227
 226 RTVPAGATTTSSATE-----TMENVKCKNFKLSTLIKLASGKQSTETANKEVQNL 279
 228 LDKAIEAEFTKRLVYELKSSPOPHLVPLUKSVALRQLLPSNSQSTQCVQO---TS 283
 280 LDKIEAEFTKRLVYELKSSPOPHLVPLUKSVALRQLLPSNSQSTQCVQO---TS 339
 284 SDVIAVCTTCTVTPVTTVSSQSEKSIIVSGAATPRTVQVTLNPLAGPVAKAGV 343
 340 Q-----ATTALTAVALVSSVORTAKTAATVTSALQPPVLSL----- 376
 344 VTLHVGPTAATGTTAGTGLQTSKPLTVSVANTVTVTSLOPEKPYVSGTAVLSLPAV 403
 377 -----TQPTQGVGKQGPPLVITQ-----QPPKP-----GALLRPQV 410
 404 TBEETSGAALCLPSPVPSFCMDHICKPYIGTPVOIKLAQPCPVLSOPACIPITGSSSKO 463
 411 TLVQT-----PMVALRQPH-NRMLTTPQOQL----- 437
 464 LFSLFHVQOPSGGNEKQVTTISHSSTLTQKGGQKTMPTVTTIIPSPAPASILKQITL 523
 438 -----NPLQPPVVKPAVL 451
 524 PGNKIL---SLQASPTQKNRIKENVTSCEPDEDINDVYTSAGVNLNEENACILATNSEL 580
 452 PGTALSAVSAVAAQAQKRLKEPGGGSPFRDDDDINDVAVMAVNSEEARILATNSEL 511
 581 VGLTIOGSCDEPFLFGALQKRLDITGKKKIDITELNSDAVNLISQATQERLGLLEKTLA 640
 512 VGLTIOGSCDEPFLFGALQKRLDITGKKKIDITELHPDVVSYVSHATQORLQNLVERISE 571
 641 IAOHRTTYKASENVYILCSPTRSOLFLEKLDLEKOROLEEREMILAKAKRSNKEDP 700
 572 TPAQKPFSTKDDRYDQASVYRAQLKFFEQLDIEROKDEERELIMRAKRSRQEDP 631
 701 EQLRLKQAKEMOQOELAQIOHRDANLTAAIGPRKKRPLE-----SGTEGKLDNLS 755

DB 632 EQLRLKQAKEMOQOELAQIOHRDANLTAAIGPRKKRKYDCPSSGAGSGSPSV 691
 QY 756 GTSSLTATFKQLHRPRITRICLDLIFCMEQEREMKYSRALYALIAK 801
 DB 692 GSSGVGTPTQFTQRTIRVNLRLDIFCLENDERETSHSLLYKAPIAK 737
 RESULT 7
 ID AAW82954 standard; Protein; 1023 AA.
 XX
 AC AAW82954;
 XX
 DT 23-APR-2002 (first entry)
 XX
 DE Human homologue of Mpr1 protein target for antifungal compound.
 XX
 KM Antifungal; fungal gene transcription; RPC34; POP3; TFA2; NAB2;
 KM MPT1; MPR2; BOS1; POL30; RSR2; SOT1; MRM1; TRB1; SPC98; BPR2; RNL1;
 KM GCD7; SK16; NIP1; LCP5; NCE103; ECOL; ORC2; CNS1; YPD1; TIM10; SRB4;
 KM Yeast; fungus.
 XX
 OS Homo sapiens.
 XX
 PN WO200202055-A2.
 XX
 PD 10-JAN-2002;
 XX
 PF 28-JUN-2001; 2001WO-US20592.
 XX
 PR 29-JUN-2000; 2000US-215164P.
 PR 10-AUG-2000; 2000US-224457P.
 XX
 PA (ANAD-) ANADYS PHARM INC.
 XX
 PI Moore J, Buurman ET, Desilva T, Harris S, Komarnitsky S;
 PI Mendillo M, Moore D, McCoy M, Sanderson K, Haq T, Zhu S, Long F;
 PI Davidov E, Thompson CM;
 XX
 DR N-PSDB; ABK32842.
 DR
 XX
 PT Screening candidate antifungal compound for interaction with essential
 PT protein, modulation of essential protein activity, binding to essential
 PT protein, by contacting protein with test compound and determining
 PT effects
 XX
 PS Claim 1; Figure 79; 522pp; English.
 CC
 CC The invention describes a method of screening a candidate antifungal
 CC compound for interaction with essential proteins (EP) or for modulation
 CC of EP activity e.g fungal gene transcription. The proteins tested in the
 CC invention include RPC34, POP3, TFA2, NAB2, MPT1, MPR2, BOS1, POL30, RSR2,
 CC SOT1, MRM1, TRB1, SPC98, BPR2, RNL1, GCD7, SK16, NIP1, LCP5, NCE103,
 CC ECOL, ORC2, CNS1, YPD1, TIM10 and SRB4 from S. cerevisiae, C. albicans
 CC and human homologues. The method involves contacting a culture with one
 CC or more test compounds and determining the effects on the growth or
 CC viability of the culture of cells which preferably comprises fungal cells
 CC or yeast cells. Preferably the identified compounds interact with, or
 CC modulate (preferably inhibit) activity of C. albicans EP. The inhibitor
 CC compounds identified by the method are useful for preventing or
 CC inhibiting fungal, particularly C. albicans growth in culture or in a
 CC mammal. The antifungal agents interact with essential fungal elements
 CC that can be used to treat fungal infection by preventing the growth and
 CC preferentially killing the fungi, but does not inhibit the biological
 CC activity of mammalian homologues. This amino acid sequence represents a
 CC target protein used to test the antifungal compounds, described in the
 CC method of the invention.
 CC
 XX
 SQ Sequence 1023 AA;
 Query Match 27 1%; Score 1157.5; DB 23; Length 1023;
 Best Local Similarity 37.6%; Pred. No. 4.8e-69;

Matches 310; Conservative 97; Mismatches 168; Indels 249; Gaps 26;

QY 13 PPKVSSG-----PRLPAQIVAVKAPNTTITQFPANLQLPQGVILIKNSGPIIM 61
 Db 414 PTATSGIRATLPTVALRPLQGP-----PQNPTNIQ---NFQLPQGMVLVRSENGQL 464
 QY 62 LVSPQGVTR-----AETSNITSRAPVANPOTVAKICVYPNSSQLIKVAVTPVKKL 115
 Db 465 MI-PQQAIAQMAQAHQAPOTTMARPRATPSAPPVQISTVQABGTPIIAR-QVTP----- 518
 QY 116 AQIGTAVTVTPKPSVQSVAVPTSVTPGKPLNTYT-TLKPSISGASS-----TPS 168
 Db 519 -----TTTIKOV---SQAKTIVQPSATLQSPGVPOLVGGAAQATASLGTAATVQGTG 571
 QY 169 NE-PNLKENSAAVOINLSPMLNKKCKNFMLIKLACSGSGSPMGONKRIYEQ 227
 Db 572 RTVGATTTSSAATE-----TMENVKCKNFELSTLIKASSGSGSTETANVELVQNL 625
 QY 228 LDAKIEBEETRKLYVELKSPQPHLYPFLKKSVALROLIPNSQSFTQOCVQOTSSDMV 287
 Db 626 LDGRIEAEDEFTSRLRIEINSSPQYLVFLKRSIPALROLTPDASAFTQSQOQPPPP-- 683
 QY 288 IATCTVTVTSPVYTTVSSQSEKSIIVSGATAPRTVSVQTLNPLAGPVAKAGVYTLH 347
 Db 684 -----TSQATTA--LTAVALSS-----SVQ-----R 702
 QY 348 SVGPATATGGTTAGTGLQTSKPLVSVANTVTVSLQPEKPVSGVAVTLSPAVTGE 407
 Db 703 TAGKTAT-----VTSALQP--PVLSLT----- 723
 QY 408 TSGAALCLPSVKPVVSCWCHICKPVIGTPOI--KLAQGPVLSQAPAGIPTSSSKOL 464
 Db 724 -----QPTQGVGVGQGOPTPL----- 739
 QY 465 FSLFHVAVQPSGNEKQVYTTISHSSTLIQKCGQKTPVNTIIPTSQPPASILK--QIT 522
 Db 740 -----VIGQPP-----KPCALIRPPQVT 757
 QY 523 LPGKILISLQASPTQKRIKENVTSCFEDDINDVTSMAGVNIENENACILATNSELVG 582
 Db 758 LTQTPVALR--QPHNRIM-----LTPQGVNLSEESARILATNSSELVG 799
 QY 583 TLVQSCDEPFLFGALQKRIIDGKHDTLMSDANVLSQATQELKRLKLAIA 642
 Db 800 TLVRSCEDEFLQAPLQRIELIGKHGTELPDVAVSVAHQOQLQVLVERISETA 859
 QY 643 QHRMTYKASENYILCSQKLEKLEKQLEKQKLEKQKLEKREMLKRAKRSNKDEPQ 702
 Db 860 QOKNESTKDDRYEQASDVRAQLKFFQDLQIEKQKDEQREKILMRKRSKQEDPEQ 919
 QY 703 LRLKQAKELQOLEIAQIQRDANLTATAIGPRKKRPLE-----SGIEGLKMLLASGT 757
 Db 920 LRLKQAKEMQOQELAQMRQDANLTALAIIGPRKKRVDGPGSGEGSGSPSVPGS 979
 QY 758 SSLTATQQLRPRITRICLDLIFCMQDEREMKSRALYIALLK 801
 Db 980 SGVGTPTQPTQRIITRVNLRLIFCLENRETSHSLLYKAFK 1023

RESULT 8
 AAR56487
 ID AAR56487 standard; Protein; 921 AA.
 AC AAR56487;
 XX
 DT 23-MAR-1995 (first entry)
 DE TATA-binding protein-associated factor dTAFl110.
 XX TATA-binding protein-associated factor dTAFl110.
 KM TATA-binding protein associated factor; dTAFl110; screening;
 XX diagnostic; therapeutic; gene transcription regulation.
 OS Drosophila.

XX XX MO9417087-A.
 PN XX
 XX 04-AUG-1994..
 PD XX
 XX 28-JAN-1994; 94WO-US01114.
 PF XX
 XX 28-JAN-1993; 93US-0013412.
 PR 30-JUN-1993; 93US-0087119.
 XX
 XX (REGC) UNIV CALIFORNIA.
 PA Comal L, Dynlacht BD, Hoey T, Rupert S, Tanese N;
 PI Tjian R, Wang E, Weinzierl ROJ;
 PI WPI: 1994-264019/32.
 DR N-PSDB: AAQ0724.
 XX
 XX TATA-binding protein associated protein factors - and
 PT corresponding nucleotide sequence and deriv. antibodies, useful
 PT in screening, diagnostics and therapeutics
 XX
 PS Disclosure; Page 56-61; 180pp; English.
 CC The TATA-binding protein associated factor dTAFl110 (including
 CC specific antibodies and fusion products) are used in drug screening,
 CC diagnostics and therapeutics. They are used in the development of
 CC specific biochemical assays for screening compounds that agonise or
 CC antagonise selected transcription factors involved in regulating
 CC gene expression associated with human pathology.
 XX
 SQ Sequence 921 AA;

Query Match 18.2%; Score 775; DB 15; Length 921;
 Best Local Similarity 29.6%; Pred. No. 2.2e-43;
 Matches 258; Conservative 126; Mismatches 308; Indels 178; Gaps 26;

QY 25 POIYAVKAPNTTITQFPANLQLPQGVILIKNSGPIIM-----VSPQOTVRAETTS 76
 Db 133 POSPSTLSTLNTGQTPA-----LLVKTQNGFQLLRVGTTGPPVTOQTITTSNNS 184
 QY 77 NITSRPANPQVYKICVTVNRSSQ-----LIKKAVTPYKLAQIGTVVTVTP 127
 Db 185 NITSSTNPTTQ--IRLQTVPAASMTNTATSNIVSVASGVANSQPHPLQOLNAQ 243
 QY 128 KPSSVQSAVPTSVYTVPGKPLNTVTLKPSLGASSTPENPLKENSAAVQINSP 187
 Db 244 APQLPQITQIOTIIPAQSQOQOVNNVSSAGGTATAVSSTA-----ATT 287
 QY 188 TMLNKK-KCKNFMLIKLACSGSGSPMGONKRIYEQLEDAKIEAEFTRKLYVELK 246
 Db 288 TQGGTKKCKRFLANLEL--STREPKVEKNVTLQELVNMNVEDEECRLERILN 345
 QY 247 SSPQPHLYPFLKKSVALROL-----LPSQSFTQ--- 276
 Db 346 ASPQCLGLFKSLPLRLQALYKELYEGIKPPQHVLAGLSQOLPKIQAIQIRIG 405
 QY 277 -----QCVQOTSSDMVIACTTCTVTTSPVYTTVSSQSEKSIIVSGATAPRTVS---VQT 329
 Db 406 PSQTTTIGQTOVRMT--TPNALGTPRTTGTHTTISKQPP--IRLPTAPRLVMTGIRT 460
 QY 330 LNLPLAGPVKAGAVVTLHSGPPTAATGTTAGTGLQTSKPLVTSVANTVTVSLQPEK 389
 Db 461 QIP-SIQVPGQANIVQIR--GQHAQLORTSGVQIRATRP-----PNSVPTAN----- 506
 QY 390 VVSGTAATVLSLPAYTFGTSQAICLPVSKPVVSCWCHICKPVYIGTPVQIKLAQGPVL 449
 Db 507 -----KLTAVKVGQTIKAI-TSLHP-----PSLAISGCP-----PPPTL 543
 QY 450 SQPAGIPGSSSK---QFSLFHVAVQPSGNEKQVYTTISHS----- 489
 Db 544 SVLSTLNSASTTTLPIPSLPTVHLPPPALRARQMONLNSHNPDAKVEIKAPSLHP 603

QY 490 -----TLTIQCGOKTNPVNTIPTSOPPPASILKQITLPKNTLSIQ-----AS 534
 Db 604 PHMERINASLTPIGAKTM-----ARPPAINKAIGKKKKDAMEMDAKLTSSGGA 654
 QY 535 PTOKNRIKENVTSCEFNDEDINDVTSAGVNLNEENACILATNSELVGTILQSCDEPFL 594
 Db 655 SAANSFEQOSSMSMYGDDDDINDVAMAGVNLAEESQRIIGC-TEINIGTOIRSCDEVEL 713
 QY 595 FIGALOKRILDIKKHIDITELNSDAVNLISQATQERLKGLEKLTIAIOHRMTTYKASEN 654
 Db 714 NIPSLQARIRATISEAGLEDPSSQDVAVLISHAQERLKNVETKAVIAEHRIIDVIKIDPR 773
 QY 655 YILCSPTRSQLEKLELDLEKORLEEREMLLKAASNSKEDDEPOLKOKAKELOQ 714
 Db 774 YEPADVROQIKFLELDKAQKHELEEREMLLRAKSSRVEDDEQAKMKARAKEMOR 833
 QY 715 LELAIOHRDANLTATATAGPRKKRPLE-----SGIEGLKDNILASGTSLSLTATKOLHRP 769
 Db 834 AEMEEELRODANLTALQALGPRKKKLKLDGETVSSGAGSSGGVLSGSAFTTL-----RP 889
 QY 770 RITRICLRDLIFCMEQEREMKYSRALYLALLK 801
 Db 890 RIKRVNLRLMFLYMEQERECRSSMLFKTYLK 921

RESULT 9

AAW06077 standard; Protein; 921 AA.

AAW06077;

27-JAN-1997 (first entry)

Drosophila TATA-binding protein associated factor dTAFII110 protein.

Drosophila; TATA-binding protein; TBP associated factor; TFIID;

RNA polymerase II; transcription; messenger RNA; nuclear fraction;

holoenzyme; lambda-gt11; expression library.

Drosophila melanogaster.

US534410-A.

09-JUL-1996.

28-JAN-1993; 93US-0013412.

28-JAN-1994; 94US-0188582.

28-JAN-1993; 93US-0013412.

30-JUN-1993; 93US-0087119.

(RBCG) UNIV CALIFORNIA.

Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;

Tjian R, Wang E, Weinzierl ROJ;

WPT; 1996-333245/33.

N-PSDB; AAT42210.

Screen for cpds. that bind human TATA-binding protein associated

factor - by testing ability to bind to polypeptide fragments of the

factor. useful as (ant)agonists of transcription factors involved in

disease.

Examples; Column 27-36; 86pp; English.

This is the amino acid sequence of the Drosophila TATA-binding protein

(TBP) associated factor (TAF) designated TAFII110. The protein is a

component of the TFIID fraction required for reconstituting RNA

polymerase II in vitro transcription activity. The encoded protein

has an estimated mol. wt. of 110 kD by SDS-PAGE, a calculated mol. wt.

based on sequence of the gene of 99.4 kD and an estimated pI of 10.1.

The invention relates to purified proteins involved in transcription

CC by RNA polymerase II, the RNA polymerase which transcribes messenger
 CC RNA. RNA polymerase II transcription proceeds in vitro upon addition
 CC of several nuclear fractions designated TFIIB, B, D, E, F, H, I and J
 CC to RNA polymerase II holoenzyme. Fraction TFIID has been shown to
 CC contain a TBP and other TAFs. Purification of TFIID and separation of
 CC its components reveals 7 proteins ranging in size from 30-250 kD.
 CC Serum raised against the TFIID fraction allowed cloning of the corresp.
 CC genes from lambda-gt11 expression libraries.
 CC XX

Sequence 921 AA:

Query Match 18.2%; Score 775; DB 17; Length 921;

Best Local Similarity 29.6%; Pred. No. 2,2e+43;

Matches 258; Conservative 128; Mismatches 308; Indels 178; Gaps 26;

25 PQLVAVKAPNTTIIOPFANQLPPGVILIKNSGPMPL-----VSPQVTRAEETS 76

Db 133 PQSPITLSLNTNGQTPA-----LVYTDNGFOLLKRGTTGPPRYVQITTNSSNS 184

QY 77 NTSRPVAPNPOTVKICIVPNSSQ-----LKKVAVTPVKLAQIGTVVTVTP 127

Db 185 NTSSTNHPTTQ-IRLQVPAASMTNTATSNIVSVASSGVANSSQPHILQDMAQ 243

QY 128 KPSSVQSAVPVSVVTPPKPLNTVYTLKPRSSIGASTPSPNEPNLKAENSAVQINLSP 187

Db 244 APQLPOTQIOTIPAOQSQOQVNNVSSAGCTAVAVSTTA-----ATT 287

QY 188 TLEENYK-KCKNFLAMLIKLAGSGSOSPEMGQVKKLVOLDLAKIEAEFTKRYELK 246

Db 288 TQGNTRKEKCKFLANIEL-STREPKPYEKVNTLQELVANNVPEEFCRLERLNL 345

QY 247 SSPQPLVPLPKSVVALROL-----LPSQSFTQ-- 276

Db 346 ASPQPLIGFLKSLPLRALYKELVIGIKPPQHVGLAGLSQQLPKIQAOIRPIG 405

QY 277 ----QCVQOSSMVIATCTTWTTSPTVTVTVSSQSEKSIIVSGATARTVS---VOT 329

Db 406 PSQTTTIGQVQVMT--TPNALGTPRTTIGTTTISKQPN---IRLPTARLNTGQIRT 460

QY 330 LNPAGVGAAGVNTLHSGPTAATGTAAGTGLQTSKPLVTSVANTVTVSLOPEK 389

Db 461 QIP-SLOVPGQANIVQIR--GPHQAQIQRGTSQVQIRATTPR-----PNSVPTAN----- 506

QY 390 VVSGTAVTSLPVTGEGTSGAALCLPSVKPVVVSFCMDHICKPVIGTPVOIKLAQCPVL 449

Db 507 -----KLTAIVKVGQTOIKAI-TPSLHP-----PSLAISGSP-----PPTPL 543

QY 450 SOPAGIPTGSSSK---QLFSLFHVVOQPSGCNEKQVTTISHSS----- 489

Db 544 SVLSTLNSASTTLPPLPSLFTVHLPPALRARQMONSLNHNHFDKAVEIKADSLHP 603

QY 490 -----TLTIQCGOKTNPVNTIPTSOPPPASILKQITLPKNTLSIQ-----AS 534

Db 604 PHMERINASLTPIGAKTM-----ARPPAINKAIGKKKKDAMEMDAKLTSSGGA 654

QY 535 PTOKNRIKENVTSCEFNDEDINDVTSAGVNLNEENACILATNSELVGTILQSCDEPFL 594

Db 655 SAANSFEQOSSMSMYGDDDDINDVAMAGVNLAEESQRIIGC-TEINIGTOIRSCDEVEL 713

QY 595 FIGALOKRILDIKKHIDITELNSDAVNLISQATQERLKGLEKLTIAIOHRMTTYKASEN 654

Db 714 NIPSLQARIRATISEAGLEDPSSQDVAVLISHAQERLKNVETKAVIAEHRIIDVIKIDPR 773

QY 655 YILCSPTRSQLEKLELDLEKORLEEREMLLKAASNSKEDDEPOLKOKAKELOQ 714

Db 774 YEPADVROQIKFLELDKAQKHELEEREMLLRAKSSRVEDDEQAKMKARAKEMOR 833

QY 715 LELAIOHRDANLTATATAGPRKKRPLE-----SGIEGLKDNILASGTSLSLTATKOLHRP 769

Db 834 AEMEEELRODANLTALQALGPRKKKLKLDGETVSSGAGSSGGVLSGSAFTTL-----RP 889

QY 770 RITRICLRDLIFCMEQEREMKYSRALYLALLK 801

Db 890 RIKRVNLDMLFYMEQERFCRSMLEFITYLK 921

RESULT 10
AAM25028

ID AAM25028 standard; Protein: 921 AA.

XX
AC AAM25028;

DT 08-OCT-1997 (first entry)

DE TATA-binding protein associated factor, dTAFII110.

XX
XX TATA-binding protein associated factor; TAF; nuclear protein;
KM RNA polymerase transcription; TATA-binding protein; TBP;
KM Initiation.

XX
OS Drosophila sp.

XX
PN US5637686-A.

PD 10-JUN-1997.

XX
XX 28-JAN-1993; 93US-0013412.

PR 28-JAN-1994; 94US-0188582.

PR 28-JAN-1993; 93US-0013412.

PR 30-JUN-1993; 93US-0087119.

PR 09-MAY-1996; 96US-0646715.

XX
PA (REGC) UNIV CALIFORNIA.

PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
PI Tjian R, Wang E, Weinzierl ROJ;
XX
DR WPI: 1997-319113/29.
XX
XX N-PSDB: AAT/9604.

XX
XX Nucleic acids encoding human TATA-binding protein associated factor
PT (TAF) peptide(s) - for production of recombinant peptide(s), used
PT for modulating transcription of TAFs

XX
XX Example 1: Column 35-40; 86bp; English.

XX
XX AAM25028 represents TATA-binding protein associated factor (TAF)
CC polypeptide, dTAFII110 (mol. weight 110kD). TAF peptides derived
CC from dTAFII110 alpha, dTAFII110 beta, dTAFII140, dTAFII60, dTAFII80,
CC dTAFII110, dTAFII150, and dTAFII250, their human equivalents and
CC nucleic acids encoding them, are used to modulate transcription,
CC including transcription initiation. TAFs are nuclear proteins involved
CC in RNA polymerase I, II and III transcription. The peptides act by
CC binding to a different TAF, an activator, or TBP (TATA-binding protein)
CC or competitively inhibiting association of a TAF domain with another
CC compound, typically a protein like TBP or another TAF, an activator,
CC or DNA.

XX
XX
XX
SQ Sequence 921 AA;

Query Match 18.28; Score 775; DB 18; Length 921;
Best Local Similarity 29.68; Pred. No. 2.2e-43;
Matches 256; Conservative 128; Mismatches 308; Indels 178; Gaps 26;

QY 25 POIYAVKAPNTTITOPANLQLPGTVLIRKNSGPLM-----VSPOOTVRAETTS 76
DB 133 PQSIFSLSTLNTGOTRA-----LLVKTDNGFOLLRVCTTGPPIVYOTITINTSNN 184
QY 77 NITSRPVPANPQTVKICTYVNSSQ-----LIKKAVTPEYKLAQIGITVVTVP 127
DB 185 NITSTNHPPTTQ-IRLQTVPAASMTWTATSNLIYNSVASSGVANSOPPHLTQLNAQ 243
QY 128 KPSSVQAVAVTSVYVTPGRLNVTYTLKRSLSGASTPSENERLKAENSAVAQINIS 187
DB 244 APQLPQITQITIPAOQSQQQVNNVSSAGGTATAVSSTTA-----ATT 287

QY 188 TMELENVK-KCKNFLAMLIKILACSGSQSPKMGONKVEQLLDKIEAEFEFRKLYVELK 246
DB 288 TQGGNTKCKCRKFLANLJEL--STREPKPVKKNVRLTLQELVYNNVNEEEFCDRLERLLN 345
QY 247 SSPQHLVPELKKSVVALRQI-----LPSQSFQ--- 276
DB 346 ASPQCLIGFLKSLIPLLRQALYKEVIEGIRKPPQHVLGAGISQQLPRQIAQIRIG 405
QY 277 ----QCVOQTSDDWAIACCTTIVTTSPPVYTTVSSQSEKSIYSGANAPRTVS---VQT 329
DB 406 PSQTTTIGOTGVRRM--TPNALGTPRPTIGHITTSKQPPN---IRLPAPRLVNTGIGRT 460
QY 330 INPLAGPVGAAVYTLHVSPTAATGCTAGTGLQTSKPLVTSVANTYTVTSIQPEKP 389
DB 461 QLP-SLQVPGGANIVQIR--GPNHQLORTGVSQVRAITRP-----PNSVTPAN----- 506
QY 390 VVSGTAVTSLPVAVTGEGTSGAALCPSPKPVVSFCMDHICKPVIGTPVQIKLAOPGVL 449
DB 507 -----KTLAVKVGQTIKAI-TPSLHP-----PSLAALISGCP-----PPTPTL 543
QY 450 SQPAGIPRTGSSSK---QLFSLFHYVQPSGNGEKQVTTISHSS----- 489
DB 544 SVLSTLNSASTTLPPLPPLVTHLPEALRAREQKQNLNNSNHFDAKVEIKAPSLHP 603
QY 490 -----TLTIQKCGQKTPVNTIIPTSQPPASILKQITLPGNKILSIQ-----AS 534
DB 604 PHMERINSLRPIGAKTM-----ARPPAIKAIKAGKKKRDAMENDAKLNTSSGGA 654
QY 535 PTQKNRIKENVTSCEFDEDDINDVTSMAGVNLNEENACILATNSELVGLTIQSCDEPFL 594
DB 655 SAANSFFQSSMSMWGDDINDVAAAGVNLAEESQRIKLC-TEINIGTQIRSCDEVEL 713
QY 595 FIGAQRIIDIGKKHDTTELNSDAVNLISQATQORLGLLEKLTALAQHRTTYKAEN 654
DB 714 NLPSLQARIRAITSEAGDEPSQDAVLAISHACERLNIYEKLVAVIEHRDIVIKLDP 773
QY 655 YILCSDTQSQKLEKLEKQKRDLEEREMILKAASRSNKEDPEQLRKOKAKEIQ 714
DB 774 YEPKADVAGQIKFLFELDKAQKREHELEREMILRAASRSRVEDPEQAKKAKKAKKQ 833
QY 715 LELAOIQRDANLATAAIGPKRRKPLE-----SGIEGLKDNLASGTSSTLATQKLRP 769
DB 834 AEMELRQRDANLFLAQIIGPKRKLKLDGETVSSGAGSSGGVLTSSGSAPPTL---RP 889
QY 770 RITRCLDLFLCMQEREMKYSRALYLALAL 801
DB 890 RIKRVNLDMLFYMEQERFCRSMLEFITYLK 921

RESULT 11
ABB61528

ID ABB61528 standard; Protein: 921 AA.

XX
AC ABB61528;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 11376.

XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical.

XX
XX Drosophila melanogaster.

XX
PN WO200171042-A2.

PD 27-SEP-2001.

XX
XX 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

Db 2227 QVO---TTTSQPI---PIQPHMSLOIPISOGQPOPOVOSSTOTLS-----SCGOTLN 2272
 QY 481 QVTTTSHS-STLTQKCGKTMPTVNTIIPTSQF-PPASILKQITLPGNKILSLQASPTOK 538
 Db 2273 QVSSVSPSRPQLOIQ-----POPOVIAPVLOQOQVQVLSQI--OSQVVAQIOA---OQ 2321
 QY 539 NRIKENV-----TSCFRDEDDINDVYSMAGVNLNEMNACILATNSELVTGLIOSCK 589
 Db 2322 SGVPQIKIQLPIQIOSSAVQTHQIQNVVTVQAAVQEQI-----QVQOQLR 2369
 QY 590 DEPFLEIGALQKR---ILDIGKKHDITELNSDAVNLIISOATQERLGLLEKTLTAIAQH-- 644
 Db 2370 DQO-----QKKQOQIIEIKRHTIQASNQ-----SELQKQ---VVKHNAVIEHLK 2413
 QY 645 ---RMTTYKASEN--YILCSDFRSQKLEKLDQLEKQKRDLEEREMLLKAASRSNKED 699
 Db 2414 QKSMTPAREERQRMIVCNQVKYI--LDKIDKEEKQ-----AAKKRRRES 2459
 QY 700 PEOLRIKOKAKELQOL 715
 Db 2460 VEOKRSKQNAKTKLSAL 2475

RESULT 14

AA57452
 ID AA57452 standard; Protein; 2907 AA.

AC AA57452;

DT 22-FEB-2000 (first entry)

DE Human transcriptional regulatory factor SEQ ID NO:1.

KM Human; transcriptional regulatory factor; TCOAL; BLAST detection;

KW bromo-domain; cell proliferation; cancer.

OS Homo sapiens.

PN WO957143-A1.

PD 11-NOV-1999.

PE 30-APR-1999; 99WO-JP02340.

PR 30-APR-1998; 98JP-0137631.

PA (CHUGAI) CHUGAI RES INST MOLECULAR MEDICINE INC.

PI Jones MH;

DR WPI: 2000-052940/04.

DR N-PSDB; AA239032.

PT Transcriptional regulatory factor containing a bromo domain and gene

TCOAL encoding it

PS Claim 1; Page 48-68; 154pp; Japanese.

CC The present sequence represents a human transcriptional regulatory factor

CC containing a bromo domain. The factor interacts with proteins involved

CC in the chromatin-mediated transcription regulatory mechanism. It binds

CC to hSNF2h, hSNF2L and NCOA-62/Skip. It can be used for screening

CC compounds binding to it and acting as agonists or antagonists, which

CC are potentially useful for the treatment and prevention of cancer and

CC other cell proliferation disorders.

XX Sequence 2907 AA;

XX Query Match 6.0%; Score 255.5; DB 21; Length 2907;

XX Best Local Similarity 22.8%; Pred. No. 1e-07;

XX Matches 181; Conservative 101; Mismatches 298; Indels 213; Gaps 37;

QY 7 VAPVSAAPRVSSGPRLPAPQIYAVKAPNTTTIQFANQLPCTVLI--KNSNGPLMLYS 64

Db 1938 VAPISG--SYTTGCKMVL--TTKVGSPATVTEQONKNEHOTFAVWAGQSSNG---VVO 1990
 QY 65 POQVTRAEETSNITSRAVPA--NPQVTKICTVNP-----SSQLIKKVAVTPVKKL 115
 Db 1991 VQOKVGLIIPSSGTISQOTTSFQPRATVYITRNPNSGSGTISNNGVITGPQIRP----- 2046
 QY 116 AQTGTVVTVTPKPSVQSAVAPPSVTVTPGKPLNTVTTL---KP--SSLGASPPSNE 170
 Db 2047 ---GNTVIRT-PLQOSTLGRKAITRPVWVQPGAPQVMTQIIRGQPVSTVASPNTVST 2102
 QY 171 PNKAEMSAVQINLSPTMLNKKCNFLAMLKILACSSGSPKEMQNKVLEQLDA 230
 Db 2103 PGKSLTSA-----TSSNTIOS-----SASQPPRQOQGVKVTMAQLTQ 2141
 QY 231 KIEAEETRLKVELKSSPPHLPFLKSSVALRQLLPSSQFI---QCVQO----- 281
 Db 2142 LTQGHGNGOGLVYVIOGQO-----TTGQLDLIPGCVTVLPEPGQLMAQAMPNG 2191
 QY 282 TSSDMVIANCTTTVTSPVTVTVSSS-----QSEKSI--TVSGATA 321
 Db 2192 TVQRFLETPLATATATASTTTTVSTTAAGTGROGSKLSPQMVHQDKTLPPAQSSVSG 2251
 QY 322 PRTVSVOTLNPLAG-----VGAKAGVTVLHSG-----PTAATGTTA 360
 Db 2252 PAKAPQTAQPSARPOQPOPOSPAPQEPVOTQEPVOTTVSSHVSEAPTHA----- 2305
 QY 361 GTGLQTSKPLVTSVANTVTVVSLQPEKPYVSGTAVTSLPAVTFGEGTGAALCLPSVKP 420
 Db 2306 ---QSKRPVAAQS-----QPSNVQGSQPVAVQSPQTRIRPSTPQSLSPQGS 2352
 QY 421 VGFECMDHICKPVYIGTPVQIKLAQPGFVLSQPAIGIPGSSSKOLFSLFHVQOPSGCNEK 480
 Db 2353 QVO---TTTSQPI---PIQPHMSLOIPISOGQPOPOVOSSTOTLS-----SCGOTLN 2398
 QY 481 QVTTTSHS-STLTQKCGKTMPTVNTIIPTSQF-PPASILKQITLPGNKILSLQASPTOK 538
 Db 2399 QVSSVSPSRPQLOIQ-----POPOVIAPVLOQOQVQVLSQI--OSQVVAQIOA---OQ 2447
 QY 539 NRIKENV-----TSCFRDEDDINDVYSMAGVNLNEMNACILATNSELVTGLIOSCK 589
 Db 2448 SGVPQIKIQLPIQIOSSAVQTHQIQNVVTVQAAVQEQI-----QVQOQLR 2495
 QY 590 DEPFLEIGALQKRILDIGKKHDITELNSDAVNLIISOATQERLGLLEKTLTAIAQH----- 644
 Db 2496 DQO-----QK-----KKQOQIENVNTPSKLLIKVEIIOQVYVVKHNAVIEHLKOKK 2542
 QY 645 RMTTYKASEN--YILCSDFRSQKLEKLDQLEKQKRDLEEREMLLKAASRSNKEDPEQ 702
 Db 2543 SMTPAERENQRMIVCNQVKYI--LDKIDKEEKQ-----AAKKRRRESVEQ 2588
 QY 703 LRLKQAKELQOL 715
 Db 2589 KRKQNAKTKLSAL 2601

RESULT 15

AA57141
 ID AA57141 standard; Protein; 2035 AA.

AC AA57141;

DT 19-MAR-1995 (first entry)

DE Host cell factor protein.

KW Herpes simplex virus; herpes virus; VP16; immediate early gene;

KW host cell factor; virus infection therapy; cellular protein;

OS Homo sapiens.

QY key Location/Qualifiers

FT Active-site 21..31
FT /note- "peptide R60"
FT 168..186
FT /note- "peptide R37"
FT 333..340
FT /note- "peptide R52"
FT 426..449
FT /note- "peptide 362"
FT 511..526
FT /note- "peptide 329"
FT 578..594
FT /note- "peptide 223 first sequence"
FT 594..611
FT /note- "peptide R26 1st peptide"
FT 611..623
FT /note- "peptide 223 2nd sequence"
FT 723..731
FT /note- "peptide 318"
FT 802..813
FT /note- "peptide 299"
FT 813..820
FT /note- "peptide 268"
FT 836..847
FT /note- "peptide R26 2nd sequence"
FT 1010..1031
FT /note- "THE TNT repeat 1"
FT 1072..1093
FT /note- "THE TNT repeat 2"
FT 1101..1126
FT /note- "THE TNT repeat 3"
FT 1158..1183
FT /note- "THE TNT repeat 4"
FT 1286..1311
FT /note- "THE TNT repeat 5"
FT 1314..1339
FT /note- "THE TNT repeat 6"
FT 1349..1374
FT /note- "THE TNT repeat 7"
FT 1414..1439
FT /note- "THE TNT repeat 8"
FT 1774..1781
FT /note- "peptide 293 2nd sequence"
FT 1808..1819
FT /note- "peptide 115"
FT 1819..1840
FT /note- "peptide 261 1st sequence"
FT 1853..1863
FT /note- "peptide 240"
FT 1901..1919
FT /note- "peptide R32"
FT 1919..1930
FT /note- "peptide 261 2nd sequence"
XX
PN MO9413315-A.
XX
PD 23-JUN-1994.
XX
PF 03-DEC-1993; 93MO-US11721.
XX
PR 04-DEC-1992; 92US-0989842.
PR 12-APR-1993; 93US-0046585.
XX
PA (COLD-) COLD SPRING HARBOR LAB.
PA (TULA-) TULARIK INC.
XX
PI Herr W, Lamarca K, Wilson A;
XX WPI; 1994-234207/28.
DR N-PSDB; AAO69229.
XX
PT New Host Cell Factor polypeptide(s) and nucleic acid - are used
PT to develop agents for diagnosis or treatment of disease
PT associated with expression of a HCF-modulated gene e.g. viral

PT infections
XX
PS Disclosure: Page 39; 71pp: English.
XX
CC HCF is required for the transcription of a number of
CC viral genes, such as the immediate early herpes simplex virus-1
CC genes. Epitopes of the encoded protein can be used in
CC defining functional domains of HCF, identifying compounds that
CC associate with HCF or designing compounds capable of modifying HCF
CC transcription. Such agents can be used to treat viral infections.
XX
SQ Sequence 2035 AA;
Query Match 5.6%; Score 237; DB 15; Length 2035;
Best Local Similarity 22.6%; Pred. No. 1,1e-06;
Matches 148; Conservative 84; Mismatches 239; Indels 184; Gaps 32;
OY 10 VSAPKVSQPRPAPQVAVKAPNTTIIOPANFOLPPGVILKNSGPMLVSPQOTV 69
DB 537 IGSSPQMSGMALAAAAATOKIPPSA---PTVLSPAGTTIVKT---NAVTPGTTT 588
OY 70 TRAEITSNITSRPAPVANDPQFKICTVNSSOLIKKVAVPVKKLAIGTTV----- 122
DB 589 LPA--TVKVVASSPVAVSNPAT-----RMKRTAA-----AOVGSVSSATNTS 628
OY 123 ---VTTVPKPSVQSVAVPTSVVTVPGKPLNTVTLK-PSSL-GASSTPSNEMPL---- 173
DB 629 TRPITTVHKSGTV-IVAAQAAVTVTVGVGVTITILVSPISVPGSALISMKGVMVY 687
OY 174 --KAENSAVQINLSPTMLENVKCKKFL--AMLIKIACSGSQSP-----EMGQNVK 221
DB 688 QTKPVQTSAVVQOASTGVTOIITQKGPLPATIILKVLTSAGCKPPTIITTTQASGAGK 747
OY 222 KLVEQLDLAKIEAEFTKRLVLELKSPPQPHVPLPKSVVALRDLNPSQFIQOCVQ 281
DB 748 PTILGI-----SSVSPSTT---KQGTITITIPMSALITTAAGATG 785
OY 282 TSSDMVIACTTTVT-----SPVVT-----TVSSQSEKSIIVSGA-----TA 321
DB 786 VTSSPGIKSPIITITKMTSGTGAPAKIITAVPKIATGHGQGVTVYVLKGAPOGPETI 845
OY 322 PRT-----VSQTLNPLAGPVGAK--AGVVTLSHSGPTAAGCTTAGTLLOTS 368
DB 846 LRTVPMGCVRLVTPVTVAVAVKPAVTVTLVKGTTGVTIGTVGVST--SLAGAGHSTIS 903
OY 369 KPLVTSVA--NTVTVLSLOPEKPVVSGTAVTLSTLPAVTFEGESGAICLPVSKPVVSCW 426
DB 904 ASLATPITTLGTLATLSSQ---VINPTAIVVSAAOITLLAAGLTPPTTMOV----- 954
OY 427 DHICKPVIGTPVQIKL-AOPGVLSQPA-GIPVGSSSKQLFSLFHVVOOPSGGNEKOVTT 484
DB 955 -----SQPTQVTLTAPSGVENQPVHDLVY-----ILNSP-----TT 987
OY 485 ISHSSTLTIOCKGQKTMVNTIIPTSOPPPASI-----LKQITLPPNKILISLQASPTQ 537
DB 988 EOPFATVITADSGQGDVOPGVTVLCSNPCEHETHTTNTATTVVAN--LGHBPQPTQ 1045
OY 538 KNRIKENTVSCFDEDDINDVTSAGVNLNENACIILATNSLVTLLQSCDEP 592
DB 1046 VQFV-----CDROEAAASLVITVG---QDN-----GSVVAVCSNPP 1079

Search completed: February 16, 2003, 21:54:38
Job time : 83.3139 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:56:34 ; Search time 16.7846 Seconds

(without alignments)
1296.883 Million cell updates/sec

Title: US-09-763-909-2

Perfect score: 4264
Sequence: 1 GILVTVAPVAPVAPKVSQGP.....KZHOHERALFTITLTLTV 852

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 140259 seqs, 25548876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB pep: *
2: /cgn2_6/ptodata/2/pubppa/PCY_NEW_PUB pep: *
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB pep: *
4: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB pep: *
5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB pep: *
6: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB pep: *
7: /cgn2_6/ptodata/2/pubppa/PCY05_PUBCOMB pep: *
8: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB pep: *
9: /cgn2_6/ptodata/2/pubppa/US09_NEW_PUB pep: *
10: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB pep: *
11: /cgn2_6/ptodata/2/pubppa/US10_NEW_PUB pep: *
12: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB pep: *
13: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB pep: *
14: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	1157.5	27.1	1023	US-09-893-519A-14	Sequence 14, App1
2	220	5.2	1367	US-09-801-368-108	Sequence 108, App
3	219	5.1	5179	US-10-025-380-1068	Sequence 1068, App
4	219	5.1	5179	US-09-922-217-1068	Sequence 1068, App
5	219	5.1	5179	US-09-833-263-1068	Sequence 1068, App
6	203.5	4.8	1332	US-09-801-368-114	Sequence 114, App
7	184.5	4.3	1337	US-09-801-368-104	Sequence 104, App
8	182.5	4.3	1601	US-09-862-027-40	Sequence 40, App
9	182	4.3	1075	US-09-801-368-110	Sequence 110, App
10	182	4.3	2665	US-09-864-761-34248	Sequence 34248, App
11	172	4.0	1169	US-09-801-368-106	Sequence 106, App
12	167	3.9	1056	US-10-161-510-10	Sequence 10, App1
13	159.5	3.7	2828	US-09-905-129-21	Sequence 21, App1
14	159.5	3.7	2828	US-09-991-630-21	Sequence 21, App1
15	158.5	3.7	2597	US-09-905-129-2	Sequence 2, App1
16	158.5	3.7	2597	US-09-905-129-10	Sequence 10, App1
17	158.5	3.7	2597	US-09-905-129-13	Sequence 13, App1
18	158.5	3.7	2597	US-09-991-630-12	Sequence 12, App1
19	158.5	3.7	2597	US-09-991-630-10	Sequence 10, App1

20	158.5	3.7	2597	10	US-09-991-630-13	Sequence 13, App1
21	157.5	3.7	941	12	US-10-124-557-14	Sequence 14, App1
22	157.5	3.7	1022	12	US-10-124-557-84	Sequence 84, App1
23	157.5	3.7	1038	12	US-10-124-557-74	Sequence 74, App1
24	157.5	3.7	1049	12	US-10-124-557-58	Sequence 58, App1
25	157.5	3.7	1140	12	US-10-124-557-104	Sequence 104, App
26	157.5	3.7	1270	12	US-10-124-557-44	Sequence 44, App1
27	157.5	3.7	1311	12	US-10-124-557-42	Sequence 42, App
28	157.5	3.7	1314	12	US-10-124-557-50	Sequence 50, App
29	157.5	3.7	1320	12	US-10-124-557-46	Sequence 46, App1
30	157.5	3.7	1320	12	US-10-124-557-60	Sequence 60, App1
31	157.5	3.7	1320	12	US-10-124-557-48	Sequence 48, App1
32	157.5	3.7	1354	12	US-10-124-557-40	Sequence 40, App1
33	157.5	3.7	1361	12	US-10-124-557-40	Sequence 40, App1
34	157.5	3.7	1363	12	US-10-124-557-52	Sequence 52, App1
35	157.5	3.7	1404	12	US-10-124-557-2	Sequence 2, App1
36	157.5	3.7	1404	12	US-10-124-557-62	Sequence 62, App1
37	157	3.7	594	10	US-09-801-368-112	Sequence 112, App
38	157	3.7	688	10	US-09-864-761-36047	Sequence 36047, A
39	156	3.7	2586	10	US-09-905-129-11	Sequence 11, App1
40	156	3.7	2586	10	US-09-905-129-14	Sequence 14, App1
41	156	3.7	2586	10	US-09-991-630-11	Sequence 11, App1
42	156	3.7	2586	10	US-09-991-630-14	Sequence 14, App1
43	156	3.7	2587	10	US-09-905-129-16	Sequence 16, App1
44	156	3.7	2587	10	US-09-991-630-16	Sequence 16, App1
45	156	3.7	2589	10	US-09-991-630-24	Sequence 24, App1

ALIGNMENTS

RESULT 1
US-09-893-519A-14
: Sequence 14, Application US/09893519A
: Publication NO. US20030027243A1
: GENERAL INFORMATION:
: APPLICANT: ANADYS PHARMACEUTICALS, INC.
: APPLICANT: THOMPSON, Craig
: APPLICANT: MOORE, Jeffrey
: APPLICANT: BUDRMAN, Ed T.
: APPLICANT: BRADLEY, John
: APPLICANT: DESILVA, Thamara
: APPLICANT: HARRIS, Sandra
: APPLICANT: KOMARNITSKY, Svetlana
: APPLICANT: MOORE, Daniel
: APPLICANT: MCCOY, Melissa
: APPLICANT: SANDERSON, Karen
: APPLICANT: HAQ, Tariq
: APPLICANT: ZHU, Shuhao
: APPLICANT: LONG, Fan
: APPLICANT: DAVIDOV, Eugene
: TITLE OR INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
: FILE REFERENCE: 0342/1G348-US2
: CURRENT APPLICATION NUMBER: US/09/893, 519A
: CURRENT FILING DATE: 2001-06-28
: PRIOR APPLICATION NUMBER: US 60/215,164
: PRIOR FILING DATE: 2000-06-29
: PRIOR APPLICATION NUMBER: US 60/224,457
: PRIOR FILING DATE: 2000-08-10
: NUMBER OF SEQ ID NOS: 146
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 14
: LENGTH: 1023
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: OTHER INFORMATION: Corresponds to SEQ ID NO: 87
: PUBLICATION INFORMATION:
: DATABASE ACCESSION NUMBER: Human Genbank/CAA72189
: DATABASE ENTRY DATE: 1997-06-25
: RELEVANT RESIDUES: (1)..(1023)


```

Db 947 SKSVITPSS--STGSGSESTGSSASSSSSSSSSPKSTYSSSLPPTSATSGEI 1004
Qy 397 TLSLPAVTEGTCGAICLPSPKPVSEFCMDHICKPVIGTPV---QIKLAOPGVLSQP 452
Db 1005 TSSLPVTTTTSDEQTLIV-----TVTSCESHVCTESISSAIVATYVSGATTEYTTW 1059
Qy 453 AGIPTGSSSKOLFSLFHVVOQPSGGENEKOVTTISHSSTLTQKC-----GQ 498
Db 1060 CPTSTTEITKQ---TTEETKQTKGTTEQTEETKQTVVTTISCSBDCSKTASPAIVST 1116
Qy 499 KTMVNTI-----IPTSQFPASILKQITLPGNKILSLQSPKCNKIKENVTCF 549
Db 1117 STATINGTTEYTWCPSTSTESKQOQTLVTVTSCSGVCSSETTSPA---IVSTATA-- 1170
Qy 550 RDEDINDVTSAGVNLNEENACILATNSELVGLIOSCKDEPLFICALQ-----600
Db 1171 ---TVNDVTVVISTWRQ-----TTMEQSVSSKMSNATSETTNTGAETTTSGAAE 1220
Qy 601 -----KRLIDGKHHDTLELNSDAVNLISQATQERL 631
Db 1221 TKTVVTSISRFNHAETQATADVIGHSSSVSVSEGTGNKSLTSSGLSTMSQ--QPRS 1278
Qy 632 RGLLEKL-TALAOHRMTYKASENTIL 657
Db 1279 TPASSWGSSTASLEISTYAGSANSLL 1305

```

RESULT 7

```

US-09-801-368-104
; Sequence 104, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Call, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amlr
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 104
; LENGTH: 1537
; TYPE: PRF
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-104

```

```

Query Match 4.3%; Score 184.5; DB 10; Length 1537;
Best Local Similarity 19.3%; Pred. No. 0.0011;
Matches 134; Conservative 104; Mismatches 257; Indels 201; Gaps 27;

```

```

Qy 50 TVL-----KSNGLMLVSPQOTVRAETTSITSRPAVNP--QTVICIVPNSSQLI 104
Db 938 TVIIRPTBGLISTTEPTGTFTSTSTEVTTITGTCQPTDETVIIRPTSE--- 993
Qy 105 KKVAVTPVKLAQIGTVVTVPKPSSVQVAVPTSVTVPGKRLN-TVTTIK-PSSLG 162
Db 994 -----GLISTTEPTGTFTSTSTEVTTITGTCQPTDETVIIRPTSEG 1039

```

```

Qy 163 ASPTSPNEP-----NLKAENSAVAOINLSPITMLENVKCKNFLMLIKLACSGSQSPEN 216
Db 1040 LVTT-TTEPTWGTFTSTSTENSTVGTGNLPD-ETVIVVKT-PTTALSSSLSSSQOI 1096
Qy 217 GONVAKLYEQLDAKTEAEETFRKLYELKSSPOPHLYPELKKSVVALROLPLNSQFTQ 276
Db 1097 TSSI-----TSSRPILRPFPSS-----NGTSVI- 1119
Qy 277 QCVQOTSSDMVIANCTTV-PTSPVTTTVSS-----SOSKSIIV-----SGA 319
Db 1120 -----SSVTSVSSVLSLFTSSPVISSVSSSTTSTISFSESSKSVIPTSSSTSGS 1173
Qy 320 TAPRTVSQTLNPLAGPVANAGVYTLHSGVTAATGTTAGTGLTQTSKPLVSVANTV 379
Db 1174 SESESTSS-----AGSVSSSSSFISSESSKSPYSS-----SPLPVATTSIQ 1215
Qy 380 TVVSLQPEKPVVSGTAVTSLCPAVTEGETSGAICLPSPKPVVSEFCMDHIC---KPVIG 435
Db 1216 ETAS-----SLPPTATTKTSEQTLV-----TVTSCESHVCTESISPAIV 1255
Qy 436 TPVOIKLAQPGVLSQAPAGIPTGSSSKOLFSLFHVVOQPSGGENEKOVTTISHSSTLTQK 495
Db 1256 STATVTV-----SGVTEYTWCPSTTEETKQTKGTTEQTEETKQTVVTTISS 1305
Qy 496 C-----GQKTMVNTI-----IPTSQFPASILKQITLPGNKILSLQ 532
Db 1306 CESDYCSKTASPAIVSTGATINGTTEYTWCPSTSTESKQOQTLVTVTSCSGVCSSET 1365
Qy 533 ASP-----TQKNRIKENVT-----SCFRDEDINDVTSAGVNLNEENACILATNSELV 581
Db 1366 ASPAIVSTATATVNDVTVVPTWRPQTANESVSSKMSNATGETTNTLAAETTTNTVAA 1425
Qy 582 GTLIOSCKDEPLFICALQKR-----ILIDGKHHDTLELNSDAVN 621
Db 1426 ETTTGALETFTVYVTVSSLSNHAETQATADVIGHSSSVSVSEGTGNKSLTSSGLS 1485
Qy 622 LISQATQERLRLLEKLTALAOHRMTYKASENTIL 657
Db 1486 TMSQOPRSTPASSWVGCGST-ASLEISTYAGSANSLL 1520

```

RESULT 8

```

US-09-862-027-40
; Sequence 40, Application US/09862027
; Patent No. US20020142428A1
; GENERAL INFORMATION:
; APPLICANT: Hodge, Martin R.
; TITLE OF INVENTION: No. US20020142428A1el Kinases and Uses Thereof
; FILE REFERENCE: 35800/234862
; CURRENT APPLICATION NUMBER: US/09/862,027
; CURRENT FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 09/345,473
; PRIOR FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40
; LENGTH: 1601
; TYPE: PRF
; ORGANISM: C. elegans
US-09-862-027-40

```

```

Query Match 4.3%; Score 182.5; DB 10; Length 1601;
Best Local Similarity 19.4%; Pred. No. 0.0016;
Matches 136; Conservative 115; Mismatches 293; Indels 157; Gaps 29;

```

```

Qy 5 TKVAVSAPPKVSSGPRLLPAPQIVAVKAPNTTITQIPANQLPQGVILIKSNGSLMVS 64
Db 514 TPIPTPATPH-SSAQOQPIPPPLSTQ---TSADIQSQAQP-----SVPTWIA 559
Qy 65 --POOTVRAETTSITSRPAVNPQVTKICTVPNSSQLIKKVAVTPVKLAQIGTVV 122
Db 560 NIPAMSPSAPQPVLS--PTSAAVPPTTMIHVKPSEIPVQNAVTT-----AAVAVANN 613

```

; ORGANISM: Saccharomyces cerevisiae
 us-09-801-368-110
 Query Match 4.3%, Score 182; DB 10; Length 1075;
 Best Local Similarity 20.9%; Pred. No. 0.00096;
 Matches 158; Conservative 95; Mismatches 279; Indels 224; Gaps 34;
 QY 2 TLTVKAVASAPPKVSSPPRLPAQIVAAKAPN-----TTRQFPANQLPPGYLIKSN 56
 DB 427 TEVTTITGTNGQPIDET-----VIVIRPTSGRLTTTTE-----PWICFTST 470
 QY 57 SGPMLV--SFOQT-----VTRAETTSNITSRPAPV---ANPQVKTCPVNSSQLI 104
 DB 471 STEMTVTGTNGQPPDETVIATIRPTSEGLISTTTEPMTGTFTSTTEVTTITGNGQPT 530
 QY 105 KKVAA---TPKKKLAQIGTIVTTPKPSQSVAPVSVTVMPGKRLN-TVTTLKSS 160
 DB 531 DETVAVITPTSE---GLITTTTEPMTGTFTSTTEMTVTGTNGQPPDETVIATIRPT 586
 QY 161 LGASSTPSENPRLKAENSAVOI-----NLSPMLLENVKKCKNKLAMLIK---LACSGS 211
 DB 587 SEGILTRTEPEWGTGFTSTSTEVTTITGNGQPT-----DETAVITPTPALISS 637
 QY 212 QSPENGVAKIVEDLLAKITAEFTKRLVYELKSSQPHVFP-----LKKSVY- 262
 DB 638 LSSSGQITTSI-----TSSRPITTPPEPANGSTVSSISVYS 674
 QY 263 -ALROLLNSQSFIOQY-----OOTSDMVIAICTTTPVTPSPVTTTVSSSQ 309
 DB 675 SSVNTSLVTSSTSSISSTSS 734
 QY 310 SEKSITVSGATAPRTVSVQTLNPLAGPVGAKAGVTVLHSGPPTATGGTTAGTGLQTSK 369
 DB 735 SSSSSISSESPKSP-TNSSSSLPV--TSATTGQETASSL-PPATYTKTSE---QTTL 784
 QY 370 PLNTSVANTVTVTSIQPKPVSGTAATVLSLPATYFGTSGAALCPBVKP----- 420
 DB 785 VTVTSCSEHCVTEST--SSAIVSTATVVS--GVTEYTTMCPSTTETTKOTKGTTEQT 840
 QY 421 -----VVSFCMDHICKPVIGTPQVQLAOPGLASQPA---GIPY--- 457
 DB 841 KGTTEQTTETTKQTTVTVTSSCESDICS-----KTAPALVSTSTAINVGTTEY 891
 QY 458 -----GSSSQQLBSLFHYVQAPSG-GNEKQVTTISHSSTLTIOKCGOKTTPVNTIPTS 510
 DB 892 TWCPISTTESAQQTTLVTVTSCESGVCSETTSPALVSTATATVNV-----DVTYVYPTW 944
 QY 511 QEPASILKQTLTPGNKLISLOASPTOKNRJKEVTSQF-----RDEDDINDVTSAG--- 563
 DB 945 R-PQTNQSVSSKKNKMSATSETTTTGAERKTAVTSSLSRFAHETQTAATDVYIGSS 1003
 QY 564 ---VNLNEENACILATNSELVGTLIQSCDEPFLGALQKRLIDIGKKHDTTELNSDAVN 621
 DB 1004 SVVSSEGTGNMSTLSSGL-STMSQOPRSTP-----ASS 1036
 QY 622 LISQATQERLGLLEKLAIAQHRATTKYKASENTYL 657
 DB 1037 MWGST-----ASLEISTYAGSANSLL 1058
 RESULT 10
 US-09-864-761-34248
 Sequence 34248, Application US/09864761
 Patent No. US2002048763A1
 GENERAL INFORMATION:
 APPLICANT: Penn, Sharon G.
 APPLICANT: Rank, David R.
 APPLICANT: Hanzel, David K.
 APPLICANT: Chen, Wensheng
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 FILE REFERENCE: Neomica-X-1
 CURRENT APPLICATION NUMBER: US/09/864,761

```
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 34248
;; LENGTH: 2665
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL034555.2
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
;; OTHER INFORMATION: EXPRESSED IN PERAL LIVER, SIGNAL = 8.9
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.8
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.2
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.5
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.1
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.3
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 7.7
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 12
;; OTHER INFORMATION: EST_HUMAN HIT: AUL17052.1, EVALUO 0.00e+00
;; OTHER INFORMATION: SWISSPROT HIT: P08640, EVALUO 3.00e-10
US-09-864-761-34248
```

```
Query Match 4.3%; Score 182; DB 10; Length 2665;
Best Local Similarity 22.2%; Pred. No. 0.0035;
Matches 133; Conservative 76; Mismatches 230; Indels 160; Gaps 29;
```

```
Qy 9 PVSAPKVVSSGRPLPAQIVAVKAPNTTIOFPANLQLPQGVILKNSGPMIAYS--PQ 66
Db 1841 PVTTP-----SDPSIPIPLPSTYA-----AKLSPPVAGSGGIPHOSEPPKVTWETTR 1887
Qy 67 QTVTRAEFTSNITSRPAPVAPNPOIVKICTVPSNSQLIKKVAVTP--VKLIAQIGTTVVT 124
Db 1888 QEERPAQSTPS-----PALPDPDKASDVDRSSSTL-RKILMDPKYVSATVSTSVTT 1939
Qy 125 TVPKRSSVQSAVAVPTSVVTTPGKPLNVTYTLTKPSSLSGASTPSNEPNIKASNAVAQIN 184
```

```
Db 1940 AIAEPVSAAPCLHEAPPPPVDSKKPLEEKTA--PPVTNNSEIQASEVILVAADEKAVAPV- 1996
Qy 185 ISPTMLENVKCKKNFLAMLIKACSGSOSPEGNQVKKVLEDLDAKIAEEFTRLKLYE 244
Db 1997 IAPKITSYISR-----MPVSIDLENSQKITLAKAPQILITLVA----- 2036
Qy 245 LKSSDQPHLVPLPKRSVVALROLPLPNSOSFIOQCVQOOTSDDMVIACTTYTTSPP--VVT 302
Db 2037 LGLVNVSLVP-----VNALKGPVKGVSVTTLKSLVSTPAGPVNVLKGPVNVLTGVVNL 2091
Qy 303 TTVSSSOSSEKSLIIVSGAIPRTVSVQTLNPLAGPVGAKAGVYTLHVGFTPAATGTTAGT 362
Db 2092 TPNVNAF-----VGTVNAAPGTV-----NAAASAVNAASATVYAGAVTAASGCGTATTT 2140
Qy 363 GLLQTSKPLVTSVANTVYTVSLQPEKPVYSGAVVLSLPVATFEGTSGAICLPPSVKPVV 422
Db 2141 GTV----- 2155
Qy 423 SPCWDHICKPVIGTPVQIKLAQPG--PVL-SOPAGIPGSSSKQLPSLFHVVOQPSGNE 479
Db 2156 -----CKQRASANENSRR-HPGSMPIYIDRRPADAGSAGLR-----VNTSEG--- 2196
Qy 480 KOYTTISHSSTLTIOKCGQKTM-PVNTTIPTISQFPFASILKQITLPGKILSLQASPTOK 538
Db 2197 -VLLSY-----GQKTEGPORISAKISQIPASAMD-----IFEQOS-VSK 2236
Qy 539 NRK-ENWTSCEPDEDIDVTSMAG-VNLNENACILATNSELVGLIOSCK-DEPFL 594
Db 2237 SQVKPDSTYA-----SQPSPKGPQAPGAYANVATHSLVLTLAQTYNAPSYISSVKADRPFL 2292
```

```
RESULT 11
US-09-801-368-106
; Sequence 106; Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Call, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Anli
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US/09/801,368
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 106
; LENGTH: 1169
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-106
```

```
Query Match 4.0%; Score 172; DB 10; Length 1169;
Best Local Similarity 21.4%; Pred. No. 0.0047;
Matches 112; Conservative 85; Mismatches 191; Indels 136; Gaps 24;
```

```
Qy 4 YTKAPVSAAPPVSSGRPLPAQIVAVKAPNTTIOF-PANLQLPQGVILKNSGPMI 62
Db 689 VTSSSVSTPIITISESSASVYTL---PSITSEFPKSTMK---TKVVISSSPTNL 740
```

QY 63 VSPQQTVAETTSNITSRAV-----PANPQVAKICVFNSSQOLIKVANTPV 112
DB 741 ITSYDTTTSKDSVGSSTSVSLSSISLPSSTYSSASBOIHRSSIVSSNGALTSFSSSTKV 800
QY 113 KK-----LAQIGTVVTVTPK-----PSSVQSVAVPTSVTVTPGPK 149
DB 801 SSSSSSSHRTSPITSSSESGIKSGVIEESTSTSSSFHETSTRTSVQJSSQVTVTSSP 860
QY 150 LNTVTTLKPSLGC---ASSTPSNEPMLKAENSAVAOINLSPWLENNKCKNLFAMLIK 205
DB 861 ISTVA---PSTGLNSQTESTNSKETMSSENSASV-----893
QY 206 IACSGOSPEMGONKKLVLEQDLDAKEASE---FTR---KLVELKSSPOHLVPELK 258
DB 894 MPSSSATSPTG-----KVTSDETSSGFSRDRRTTVYRMTSETPSTN---EQ 936
QY 259 KSVVALROLPNSSQFIQOCVQOQSSDMVATCTTV-----TV---SPVTTTVSS 307
DB 937 TLLITVSSCSNCS-----NTVSSAVVSTATTINGITTEYTCPLSATLTVSK 989
QY 308 SOS-EKSIIVSGATPPTVVOQLNP--LAGPVGAKAGVYTLHSVGPTAATGCTAGTGL 364
DB .990 LEEBEKTLITVYSCESGVSETPASPAIVSTAVTVVYSTWSPQATNKLAVSSD- 1048
QY 365 LOTSPLVTSVANTVTVYSLQPEKPVYSGTAVTSLSLPAVTEGESSAICLPYKPVVSF 424
DB 1049 IENSASASVSEAEETKTSIRNNNEVP-TSGTTSIETHP-TTTSNSENSDNVS----- 1101
QY 425 CMHDICKPVIGTPVQIKLAOPGVLISQAPICPT---GSSSKQI 464
DB 1102 ASEAVSSKSVTNPVLSVSQ-----QPRGTASSMIGSSSTASL 1139

RESULT 12
US-10-161-510-10
; Sequence 10, Application US/10161510
; Publication No. US20020192695A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: PIBS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-074C
; CURRENT APPLICATION NUMBER: US/10/161,510
; PRIOR FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 60/296,076
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,605
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/357,253
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 1056
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-161-510-10

Query Match 3.9%; Score 167; DB 9; Length 1056;
Best Local Similarity 21.4%; Pred. No. 0.0086;
Matches 134; Conservative 61; Mismatches 238; Indels 192; Gaps 24;
QY 7 VAVVSAAPKVSGBRLPAPDIVAKAANTTIOFPANLOLPPGTVLIXNSGP-LMLVSP 65
DB 39 VAOGTGAPSKVDSSFOLEPAKKNAAL-GPSEPRL---ALAPVGPRAMASSEGPPLALASP 94
QY 66 QQTVAETTSNITSRP--AVPANPQVAKICVFNSSQOLIKKAAVVPYKLAOIGTTV 123
DB 95 -----RRIALPCTPEGOKATATHRSS-----LAPISVOL--VMSASA 132
QY 124 TTVPKSSVQSVAVPTSVTVTPGK--PLNTVTTLKPSLIGASSTPSNEPMLKAENSAV 181
DB 133 GPKPPTATGSLAPTSIGLIMPASAGPSPVYTLGPNLAPTSRDQOEP-----PA 184

QY 182 QINLSPMLENNKCKNLFAMLITLACSGOSPEMGONKKLVLEQDLDAKEAEETRKL 241
DB 185 SVGPKPPL-----AASGLSLALASEROP-----208
QY 242 YVELKSSPOHLVPELKSVVALNQLLPNSQFIQOCVQOQSSDMVATCTTVTTSVPV 301
DB 209 --ELPSTPSP--VP-----SPVL 222
QY 302 TTVVSSQSEKSIIVSGATAPRTVSVQTLNPLAGPVGAKAGVYTLHSVGPTAATGCTTAG 361
DB 223 SPQEQALAPASFTASGAASVQGTARKRDAAPARPLPASG-----HLQEPAQISGPTGS 277
QY 362 TGLIQTGK--LYTSVANTVTVYSLQPEKPVYSTANTVLSL-----P 401
DB 278 PPCIQTSPDPLPSFSFRAPREALHSSPEDDVLPRPQTLPLIDVQOGSPSEGTSPGLLSP 337
QY 402 AVTFEGTSGAICLPYKPVYSCFMDHICKPVIGTPVQIKLAOPGVLISOP-AGIPGSS 460
DB 338 TFRGASGQGVPPPLKPRPS-----PSRPSHSIPNRSFCVPAPDMALPRLGTOSTGP 392
QY 461 SKOLFSLFHVYQOPSNGENKOVTTISHSTLTIOKCGOKTAPVNTTIPTSOFPASILKQ 520
DB 393 GRCLSPMLQAEAPA-----PVTTSSSTSLSSSPWSAQ-----PTMKSDPGFRITV 439
QY 521 IT-----LEGNKLTLSQASPTQKNRIKENYVSCFRDEDDINDVTSAGVNLNENACIL 574
DB 440 VTNWVGTPAMPDDVTSL-----LHLGGDDSDGADMAIGLQEVNSML- 482
QY 575 ATNSELVGTILQSCDEPLFICAL 599
DB 483 --MKRLDALFTQWSE--LFMDAL 503

RESULT 13
US-09-905-129-21
; Sequence 21, Application US/09905129
; Patent No. US20020137705A1
; GENERAL INFORMATION:
; APPLICANT: Elnat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS TH
; FILE REFERENCE: 540579-2007.2
; CURRENT APPLICATION NUMBER: US/09/905,129
; CURRENT FILING DATE: 2001-07-13
; PRIOR FILING DATE: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/207,821
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/084,944
; PRIOR FILING DATE: 1998-05-11
; PRIOR APPLICATION NUMBER: 60/085,673
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 2828
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-905-129-21

Query Match 3.7%; Score 159.5; DB 10; Length 2828;
Best Local Similarity 20.4%; Pred. No. 0.1; Length 2828;
Matches 144; Conservative 77; Mismatches 241; Indels 245; Gaps 30;
QY 50 TVLIKNSGGLMLVSPQQTVAETTSNITSRAVAVPANPQVAKICVFN-----SSQL 103
DB 1053 TLLIKKMKMSQTLQGNLLEGDPTHSSE-----SEGESSTILPDSLTGIMSSMP 1108
QY 104 IKKAAVTPYKLAOIGTTVTVTPKPSVQSVAVPTSVTVTPGK-----148
DB 1109 VKKPAETTVGTLDDKDTTVTTTPR-----QKVA--PSTMTSTHRSRRPNRRRLRPKKR 1163
QY 149 ---PLNTVTTLKPSLIGASSTPSNEPMLKAENSAVAOINLSP-----MLENPK 194

Db 1164 HRHKQPTTFAPSET-FSTQPTQAPDIKI--SSQVSSSLVPTAMVNDVNTPRKOLEMEK 1220
QY 195 KCR-----NFLAMLIKLASGSGSPENGQNVK-----LVEQ 226
Db 1221 MAEPTSGKTRPKRKHGRKPNKHRTPTSVSSRASGSKSPSPENKHRIIVTPSSFTILLPR 1280
QY 227 LLDKRIEAE-----EFTKRLVELKSSPO-PHLVFLKSSVVALROLLPNSQSFIOQC 278
Db 1281 TVSLKTGPPYDSDLYMTTKRIY---SSYPKVOETLPLVTKPPTSDGKEI---KDVATN 1333
QY 279 VQOQSSMVA--TCCTVTVTVSPVTVTVSSSQSEKSIYSGATAPRTVSQITLNLAPL 336
Db 1334 VDKHRSIDLVTGESITNAIPTRSLVSTGMEFEKES--P 1371
QY 337 VGAKAGVVLHSGVPTAATGTTAGTGLQTSKPLTVSVANTVTVSLQPEKPVSGTAV 396
Db 1372 VGFP-----GTPWNSKRTAQPGRLQTDIPVTTSGEN-LTDPPLKLELEVDFTSE 1421
QY 397 TSLPVAVT---FGETSGAALCPSPKPVV-----SFCWDHI-----CKPVIG 435
Db 1422 FLSSLVSTFPHQEAGSSFTLSIKVEVASSQAETTTLLDQDHEETVAILLSETRQNH 1481
QY 436 TPVQIKLAQCG-----PYLSQPAIGPFGSSSKQLFSLFHV----- 470
Db 1482 TPTAARKKEPSSPSTILMSLGQTTTKPALPSPRISOASRDSKENVFLNYGNPETEA 1541
QY 471 -----VOOPSGNE-----KQVTTISHSFTLQKQOKTM----- 501
Db 1542 TPVNNECTQHMSPNEIESTSSRDAPNLSTKLEKQVGSKSLPFGPSQRODGRVHA 1601
QY 502 -----PVNTIIP-----TSQPP-----PASIL---KQ 520
Db 1602 SHQLTRVAPARILPTAVRLPEMSTQASRYFVTSQSPRHMTNKPEITTPSGALPENKQ 1661
QY 521 ITLPGNKILSLQASPTQKNRIKENVTSQFRDE--DDINDVTSNAGVN 565
Db 1662 FTTPR---LSSTTIPPLPLHMSKPSIPSKFTDRIDQNGSKYGVGN 1705

RESULT 14
US-09-991-630-21
Sequence 21, Application US/09991630
Patent No. US2002015151A1
GENERAL INFORMATION:
APPLICANT: Elnat, et al
TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THEREOF
FILE REFERENCE: 540579-2007.3
CURRENT APPLICATION NUMBER: US/09/991,630
PRIOR FILING DATE: 2001-11-06
PRIOR APPLICATION NUMBER: 09/905,129
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: 09/802,318
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: 09/729,485
PRIOR FILING DATE: 2000-12-04
NUMBER OF SEQ ID NOS: 28
SOFTWARE: PatentIn version 3.0
SEQ ID NO 21
LENGTH: 2828
TYPE: PRT
ORGANISM: homo sapiens
US-09-991-630-21

Query Match 3.7%; Score 159.5; DB 10; Length 2828;
Best Local Similarity 20.4%; Pred. No. 0.1;
Matches 144; Conservative 77; Mismatches 241; Indels 245; Gaps 30;

QY 50 TVLIKSNCSGLMLVSPQQTVAETTSNISRPAVPANPQTVKICVTPN-----SSQL 103
Db 1053 TLIIKGMKMKMSQTLQGNMLBGPDIHSRSSE-----SEGESKSTIPLDSTLIGMSSMSP 1108

QY 104 IKKVAIVPVKKLAQIGTIVTVTPKPSVSVAVPTSVTVTPGK----- 148
Db 1109 VKKPAETTVGTLDKDQDVTTVTPR-----QKVA-PSSTMSTHPRRRPNGRRLRPKFR 1163
QY 149 ---PLNTVTVLPKSLASSTPSPNEPULKAENGAAYINLSPT-----MLENVK 194
Db 1164 HRHKQPTTFAPSET-FSTQPTQAPDIKI--SSQVSSSLVPTAMVNDVNTPRKOLEMEK 1220
QY 195 KCR-----NFLAMLIKLASGSGSPENGQNVK-----LVEQ 226
Db 1221 MAEPTSGKTRPKRKHGRKPNKHRTPTSVSSRASGSKSPSPENKHRIIVTPSSFTILLPR 1280
QY 227 LLDKRIEAE-----EFTKRLVELKSSPO-PHLVFLKSSVVALROLLPNSQSFIOQC 278
Db 1281 TVSLKTGPPYDSDLYMTTKRIY---SSYPKVOETLPLVTKPPTSDGKEI---KDVATN 1333
QY 279 VQOQSSMVA--TCCTVTVTVSPVTVTVSSSQSEKSIYSGATAPRTVSQITLNLAPL 336
Db 1334 VDKHRSIDLVTGESITNAIPTRSLVSTGMEFEKES--P 1371
QY 337 VGAKAGVVLHSGVPTAATGTTAGTGLQTSKPLTVSVANTVTVSLQPEKPVSGTAV 396
Db 1372 VGFP-----GTPWNSKRTAQPGRLQTDIPVTTSGEN-LTDPPLKLELEVDFTSE 1421
QY 397 TSLPVAVT---FGETSGAALCPSPKPVV-----SFCWDHI-----CKPVIG 435
Db 1422 FLSSLVSTFPHQEAGSSFTLSIKVEVASSQAETTTLLDQDHEETVAILLSETRQNH 1481
QY 436 TPVQIKLAQCG-----PYLSQPAIGPFGSSSKQLFSLFHV----- 470
Db 1482 TPTAARKKEPSSPSTILMSLGQTTTKPALPSPRISOASRDSKENVFLNYGNPETEA 1541
QY 471 -----VOOPSGNE-----KQVTTISHSFTLQKQOKTM----- 501
Db 1542 TPVNNECTQHMSPNEIESTSSRDAPNLSTKLEKQVGSKSLPFGPSQRODGRVHA 1601
QY 502 -----PVNTIIP-----TSQPP-----PASIL---KQ 520
Db 1602 SHQLTRVAPARILPTAVRLPEMSTQASRYFVTSQSPRHMTNKPEITTPSGALPENKQ 1661
QY 521 ITLPGNKILSLQASPTQKNRIKENVTSQFRDE--DDINDVTSNAGVN 565
Db 1662 FTTPR---LSSTTIPPLPLHMSKPSIPSKFTDRIDQNGSKYGVGN 1705

RESULT 15
US-09-905-129-2
Sequence 2, Application US/09905129
Patent No. US20020137705A1
GENERAL INFORMATION:
APPLICANT: Elnat, et al
TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THEREOF
FILE REFERENCE: 540579-2007.2
CURRENT APPLICATION NUMBER: US/09/905,129
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: 09/802,318
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: 60/207,821
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: 60/084,944
PRIOR FILING DATE: 1998-05-11
PRIOR APPLICATION NUMBER: 60/085,673
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 2597
TYPE: PRT
ORGANISM: rattus species
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)..(2597)

OTHER INFORMATION: 'x' can be any amino acid
US-09-905-129-2

Query Match 3.7%; Score 158.5; DB 10; Length 2597;
Best Local Similarity 20.5%; Pred. No. 0.11;
Matches 143; Conservative 81; Mismatches 242; Indels 231; Gaps 31;

```
OY 2 TLVTKVAPYSAP--PRVSGPRLPAPQIYAVKAPNTTIQF-----PANKQLPFGI----- 50
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 932 SFTSKASTVLQIPINIPESYGPQIP---ITGVSRPSSSDISHTTADPFSFSSHPGSHHTTA 988
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 51 -----VLKNSGPIML---VSPOQTVTRAFTTSN-----ITSRPAY--PAMP 88
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 969 SLEFHIPRNNNTGNPPLSHLGRERTISGRVKKNPHRTVPLRRHRKRTVRLPAIKGPANK 1048
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 89 QTVKI-----CTVPNSSSQLIKRV-----AVTPVKK- 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1049 NVSOVPATEYPCMGCHTCSAEGLVTATLAVSSSHSLPPTNNNGVIAESTIYVKKP 1108
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 115 -----LAQIGTVVTVYVK-----PSSVQSAVPTSVV----- 142
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1109 LLLFKDKONVDIEITTTTKYSGESNHVLPTEASMTSAPTGSVSLGKSPVDSGHLSPG 1168
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 143 TVTPGKPLNTVTLKPSLSGASTSPSNEPNTLAENSAVAQINLSPTMLENVK----- 195
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1169 TTQTKGD-SVETTPLPSPSTPSIPTSTKSKRKTIP-----LHQITVNNQKKEGMLKNP 1221
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 196 ----CKNFLAMLIKLA---CSGSQSPENGQNVKKLVEQLDAKIEAEFTKRLYVELKS 247
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1222 YQFGLQKNPAKLPKIPAPLLPTGQSSPS-----DS-----TTLT 1256
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 248 SPQPHL-----VPFLKKSVALROLPLPNSQFTQOCVQOTSSDMVIATCTT 293
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1257 SPPEALSTTMAATONKGTENVSGARSLGAKKQPTNNSPVLPTISKRSNTLNFLESTET 1316
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 294 TVTTPPVVTVTVSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAK--AGVVTLSHSGP 351
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1317 PTVTSPTATASVIMSETQ-----RTSKKAKDQIKGPRKRNANNTTPROVSGY 1365
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 352 TAATGGTAGTGLQTSKPLVTSVANVTVTVSLOPEKPVSGTAV--TLSLPAYT----- 404
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1366 SAYSALTATADTFLAFSHSPRODDGN-----VSAVAVHSTSLAITELEFEK 1412
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 405 FGETSGAALICLPVSVPVVSFCDHICKPVYIGTPVQIKLAQPGPVLSQPA-GIPTGS---- 459
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1413 YTOTLGNNTTALLETTL-----LSKSOESTTVKRASDTPPPLSSGAPVPPTSPPPF 1463
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 460 -----SSKOLFELFHYOQPPSGNGEKOVTTISHS--TLTIQKCGQKTMVNTII----- 507
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1464 TKGVYTDSDKVTSAFQM-----TSNRVVTIYESSRHNHTDLOOPSNAEASNPETITGTTD 1516
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 508 -PTSQFPAPASI-LKQITLPGNKILSLQASPTQKNRIK 542
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1517 SPSNLFPSISVPALRVDPQNSKMKRSPWPEHKYQLK 1553
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: February 16, 2003, 22:01:37
Job time : 60.7846 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:55:19 ; Search time 28.8257 Seconds
(without alignments)
2841.441 Million cell updates/sec

Title: US-09-763-909-2

Perfect score: 4264
Sequence: 1 GLVTRVAPVAPSVAPKVSQGP.....KZHOHERALPTIRLTLLTY 852

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR_73:.*
2: PIR1:.*
3: PIR2:.*
4: PIR3:.*
5: PIR4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	775	18.2	921	2 A48184	transcription init
2	775	18.2	921	2 A45183	TRP-associated fac
3	344	8.1	549	2 C87719	protein R119.6 [tm
4	244.5	5.7	2187	2 T30826	nascent polypeptid
5	237	5.6	2035	2 A40718	host cell factor C
6	234.5	5.5	3020	2 A43932	mucin 2 precursor,
7	233.5	5.5	528	2 I47141	gastric mucin (clo
8	230.5	5.4	1275	2 T33369	hypothetical prote
9	220	5.2	1367	2 A48478	glucan 1,4-alpha-g
10	213.5	5.0	5376	2 T42215	zonadhesin - mouse
11	212.5	5.0	2232	2 T34434	hypothetical prote
12	210.5	4.9	1777	2 T34369	hypothetical prote
13	210	4.9	873	2 A47283	calphostin - fruit
14	206.5	4.8	865	2 A47282	calcium-binding pr
15	206	4.8	1032	2 T34433	hypothetical prote
16	205.5	4.8	1161	2 S57180	probable membrane
17	204.5	4.8	1324	2 S52863	DNA-binding protei
18	203.5	4.8	873	2 F96615	probable MyD-famil
19	203	4.8	697	2 A86402	protein T22C5.17 [
20	201	4.7	725	2 A41258	a-agglutinin core
21	200.5	4.7	1367	2 S51959	hypothetical prote
22	200	4.7	3507	2 T34513	hypothetical prote
23	198	4.6	867	2 T43463	membrane glycoprot
24	197.5	4.6	881	2 S56032	probable membrane
25	197.5	4.6	1609	2 T52345	mucin 3 - human (f
26	197	4.6	648	2 PC4395	hypothetical prote
27	193.5	4.5	670	2 T36791	hypothetical prote
28	193	4.5	825	2 T29634	hypothetical prote
29	192	4.5	1229	2 T23697	hypothetical prote

30	192	4.5	3759	2 A35085	titlinhorax protein
31	191	4.5	797	1 VGBEX1	glycoprotein X pre
32	190.5	4.5	1260	2 S60896	agglutinin-like pr
33	189.5	4.4	851	2 T22696	hypothetical prote
34	189.5	4.4	866	2 T45462	membrane glycoprot
35	189.5	4.4	1630	2 A53577	asclases slaloglyco
36	188	4.4	3570	2 T45035	mucin MUC5B, trach
37	186	4.4	1151	2 T18535	high molecular mas
38	184.5	4.3	1537	2 S53465	flocculation prote
39	182.5	4.3	1353	2 T19157	probable metal bin
40	182.5	4.3	1424	2 T19156	probable metal bin
41	182.5	4.3	1851	2 T19964	hypothetical prote
42	182	4.3	1075	2 S48992	flocculation prote
43	182	4.3	3869	2 A48205	All-1 protein +GTE
44	181.5	4.3	662	2 A45155	mucin FIM-C.1 - Af
45	181.5	4.3	1199	2 A40670	nuclear envelope p

ALIGNMENTS

RESULT 1

A48184 Transcription Initiation factor IID 110k chain - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999

C:Accession: A48184

R:Kokubo, T.; Gong, D.W.; Roeder, R.G.; Horikoshi, M.; Nakatani, Y.

A:Title: The Drosophila 110-kDa transcription factor TFIID subunit directly interacts

A:Reference number: A48184; MUID:93317591; PMID:8327460

A:Accession: A48184

A:Status: preliminary

A:Molecule type: mRNA; protein

A:Residues: 1-921 <KOK>

A:Cross-references: GB:S63550; NID:9398432; PID:9398433

A:Experimental source: embryo nuclear extract

A>Note: sequence extracted from NCBI backbone (NCBI:134863, NCBI:134864)

C:Genetics:

A:Gene: FlyBase:Tafl10

A:Cross-references: FlyBase:FBgn0010280

C:Keywords: transcription Initiation

Query Match	18.2%	Score 775;	DB 2;	Length 921;
Best Local Similarity	29.6%	Pred. No. 2,1e-33;		
Matches	258;	Conservative 128;	Mismatches 308;	Indels 178; Gaps 26;
QY	25	POIYAVKAPNTTIOFPANLQLPETVLIKNSGFLM-----VSPQOTVRAETTS	76	
DB	133	POSPSITLSTLNTGOTPA-----LLVKTDCNGFQLRVGTTGPPVYOTITWTSNNS	184	
QY	77	NITSPPAVPANQPVKICVTPNSSQ-----LIKKVAVTPVKTKLAQIOCTVTVTVTP	127	
DB	185	NITSTTNHPTTTO-IRLDTPVPAASMTNTTATSNITVSVASSGANSOPHPLTOLNAQ	243	
QY	128	KPSSQVAVPVSVTVTGPKPLNTVTLKPSLSSGASTSPSEPMKAENSAVQINLSP	187	
DB	244	APQLPQIQIQITIPAQSGQQQVNVSSAGATAVSSTTA-----AFT	287	
QY	188	TMLENVK-KCKNFLMLIKLACSGSQSPKMGONKVKLVEQLIDAKIEAEFTKLYVEK	246	
DB	288	TOOGTKKCKRFLANLIEL--STREPRPVERKNVNTLQELVANAVEPEEFCDRLERLN	345	
QY	247	SSPOHLVPLKSSVVALROU-----LPSQSTIO---	276	
DB	346	ASPOCLIGFLKSLPLRLQALYKELVIEGKPPQVHLGSLQOLPKIAQIADIRIG	405	
QY	277	---OCVOOTSDMWIATCTTVTTSPPVTTVSSSQSEKSIIVSGATAPRTVS---VQT	329	
DB	406	PSQTTTIGOTQVBMF--TPNALGTPRPIIGHITISKQPPN---ILPLPAPRVNNGGIRT	460	
QY	330	LNPLAGPVGAKAGVTVLHSGFTAAATGCTAGTGLQTSKPLVTSVANTVTVTSIQPKP	389	

```

Db 461 QIP-SLOVPOQANIYQIR--GPOHAQLOQRTGSVOIRATTPR-----PNSVPTAN-----506
QY 390 VSGTAVTSLPAVTEGETSGAAILCPSPKPVVSPFCWDHICKPVIGTPVOIKLAQPCPV 449
Db 507 -----KLTAVKVGQGIKAI--PISLHP-----PSLAISGCP-----PPTPTL 543
QY 450 SOPAGIPTGSSK---QLFSLFHYVOQPSGNGEKQVTTISHS-----489
Db 544 SVLTLSNASTTTLPIPSLPTVHLPPALRARQOMNSLNHNSHFDKLVETIKASLHP 603
QY 490 -----TLTIQKCGKTMVNTIIPTSQPPASILKQITLPGNKILSLQ-----AS 534
Db 604 PHMERINASTLPIGAKTM-----ARPPAINKAIGKKKRDAMEMDAKLTSSGGA 654
QY 535 PTOKNRIVENVTSCFDEDDINDVTSNAGVNLNEENACILATNSLVLTLLIQSKDEPFL 594
Db 655 SAANSFQOOSMSMYGDDDDINDVAAAGVNLAEESORILGC--TENIGTQIRSCKEVFL 713
QY 595 FIGALQKRIIDIGKHHIDITELNSDAVNLISQATQOERLKGLEKTLTAIOHRMTTKYASEN 654
Db 714 NLPSLQARIARITSEAGLDEPSODVAVLISHAOERLKNIVERKLAIVAEHRIDIVIKIDPR 773
QY 655 YILCSPTRSQLEKLEKIDOLEKORKDEEREMLLKAKSRNKEDEPQOLKOKAKELOQ 714
Db 774 YEPAKDVROGIKFLEELDKAEQKRHELEEREMLLBAKSRVEDPEQAKMKARAKEMOR 833
QY 715 LELAQIOHRDANTATPAIGPRKKRPLE-----SGIEGLDNLLASGTSSTLTATKOLHRP 769
Db 834 AEMBELRORDANTLALQAIIGPRKKLKDGETVSSGAGSSGGVLSGSGAPVTL---RP 889
QY 770 RIRICLRDLIFCMEQEREMKYRSLAYLALLK 801
Db 890 RIKRVNLRDMLFYMEQERECRSSMLFKTYLK 921

```

RESULT 2

A:Species: *Drosophila* sp.
C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 16-Feb-1997
C:Accession: A45183
R:Hoey, T.; Wehnizler, R.O.; Gili, G.; Chen, J.L.; Dynlacht, B.D.; Tjian, R.
Cell 72, 247-260, 1993

A:Title: Molecular cloning and functional analysis of *Drosophila* TAF110 reveal properties
A:Reference number: A45183; MID:93145326; PMID:7678780
A:Accession: A45183
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-921 <HO>
A:Experimental source: embryo
A>Note: sequence extracted from NCBI backbone (NCBIP:123832)
C:Genetics:
A:Gene: FlyBase:Taf110
A:Cross-references: FlyBase:FBgn0010280

Query Match 18.2%; Score 775; DB 2; Length 921;
Best Local Similarity 29.6%; Pred. No. 2, 1e-33;
Matches 258; Conservative 128; Mismatches 308; Indels 178; Gaps 26;

```

QY 25 POIVAVKABNTTITOPANLQLEPPGTVLTKSNGSPLML-----VSPQQTVAETTS 76
Db 133 PQSPSTLSTLMTGQTPA-----LVKTDGPFLLAVGTTGTPPTVQTITNTSNN 184
QY 77 NITSRPAVNPQTVKICIVPNSSQ-----LIKKVAVTPVKLAQIGTTVTVTP 127
Db 185 NITSTNHNHTTQ--IRLQVTPAASMTNTTATSNIIIVNSVASSGVANSSQPHLQLQNAQ 243
QY 128 KPSVQSVAVPSPVTVTGKPLNTYTTILKPSLSGASSTPSNPNKAENSAVQINLSP 187
Db 244 AQLQPIQITQITPAQOQOQOQNNVSSAGTATAVSSTTA-----ATT 287
QY 188 TMLLENK--KCKNFILMLIKACSGSQSPKMGONVKVLEOLLDAKTEAEFEFTKLYVELK 246

```

```

Db 288 TQOGNTEKCKRKLANIEL--STREPKVEKENVRTLLQELVANNNEEFCORLERLNL 345
QY 247 SSPDPLVPLPKRSVVALROL-----LPSNSFQI---276
Db 346 ASQPCILIGFLKKSLLPRLQALYTKELVIGIKPPQPHVIGLAGLSQQLPKIAQIRPIG 405
QY 277 ----QCVQOTSSDMVATCTTYYTTSPTVTVSSSOSEKSIISGATAPTVS---VOT 339
Db 406 PSQTTTIGQOVNMI--TPNALGTPREPTIHTTISKPPN---IRLPTADRLVNTGIRP 460
QY 330 LNPAGPVGAKAGVNTLHSGTPAATGCTAGAGLQTSPLVTSVANTVTVSLOPEK 369
Db 461 QIP-SLOVPOQANIYQIR--GPOHAQLOQRTGSVOIRATTPR-----PNSVPTAN-----506
QY 390 VSGTAVTSLPAVTEGETSGAAILCPSPKPVVSPFCWDHICKPVIGTPVOIKLAQPCPV 449
Db 507 -----KLTAVKVGQGIKAI--PISLHP-----PSLAISGCP-----PPTPTL 543
QY 450 SOPAGIPTGSSK---QLFSLFHYVOQPSGNGEKQVTTISHS-----489
Db 544 SVLTLSNASTTTLPIPSLPTVHLPPALRARQOMNSLNHNSHFDKLVETIKASLHP 603
QY 490 -----TLTIQKCGKTMVNTIIPTSQPPASILKQITLPGNKILSLQ-----AS 534
Db 604 PHMERINASTLPIGAKTM-----ARPPAINKAIGKKKRDAMEMDAKLTSSGGA 654
QY 535 PTOKNRIVENVTSCFDEDDINDVTSNAGVNLNEENACILATNSLVLTLLIQSKDEPFL 594
Db 655 SAANSFQOOSMSMYGDDDDINDVAAAGVNLAEESORILGC--TENIGTQIRSCKEVFL 713
QY 595 FIGALQKRIIDIGKHHIDITELNSDAVNLISQATQOERLKGLEKTLTAIOHRMTTKYASEN 654
Db 714 NLPSLQARIARITSEAGLDEPSODVAVLISHAOERLKNIVERKLAIVAEHRIDIVIKIDPR 773
QY 655 YILCSPTRSQLEKLEKIDOLEKORKDEEREMLLKAKSRNKEDEPQOLKOKAKELOQ 714
Db 774 YEPAKDVROGIKFLEELDKAEQKRHELEEREMLLBAKSRVEDPEQAKMKARAKEMOR 833
QY 715 LELAQIOHRDANTATPAIGPRKKRPLE-----SGIEGLDNLLASGTSSTLTATKOLHRP 769
Db 834 AEMBELRORDANTLALQAIIGPRKKLKDGETVSSGAGSSGGVLSGSGAPVTL---RP 889
QY 770 RIRICLRDLIFCMEQEREMKYRSLAYLALLK 801
Db 890 RIKRVNLRDMLFYMEQERECRSSMLFKTYLK 921

```

RESULT 3

C:Species: *Caenorhabditis elegans*
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: C87719
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode *C. elegans*: a platform for investigating bio
A:Reference number: A75000; MID:99069613; PMID:9851916
A>Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C-
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: C87719
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-549 <STO>
A:Cross-references: GB:chr_1; PIDN:AA016427.1; PID:g3133007; GSPDB:GNO00019; CESP:R119
C:Genetics:
A:Gene: R119.6
A:Map position: 1

Query Match 8.1%; Score 344; DB 2; Length 549;
Best Local Similarity 22.7%; Pred. No. 6, 9e-11;
Matches 145; Conservative 77; Mismatches 193; Indels 224; Gaps 20;

```

QY 158 PSSLGASSTPSNPNKAENSAVQINLSPTMLLENVKKCKNFILMLIKACSGSQSPK 216

```


Db 1229 SGSTVTTGSGSSSSATSLSSSPVSPSTQSPNPSTGSSSTPTPNPSTQSTPVVS 1288
Qy 451 QPAGITGSSSKQLFS-LRHVVOQP---SGGNKQVTTTSHSLTLIQCGQRTMPVN-T 505
Db 1289 TTTGEMTSHSGSTQTPSTISTYQTPSTVSGNSGSGTIVIGSEASTSGSSFKTSPSSIS 1348
Qy 506 ITPTSQPPASILKQITLPGNKILSLQASPT 536
Db 1349 PVPSTSPISPTFASST-SGSTISDVSVST 1378

RESULT 12

t34369

hypothetical protein t19d12.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34369

R:Faustino, A.

submitted to the EMBL Data Library, November 1995

A:Description: The sequence of C. elegans cosmid t19d12.

A:Reference number: 221513

A:Accession: T34369

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1777 <FAV>

A:Cross-references: EMBL:U41263; PIDN:AAC24428.1; GSPDB:GN00020; CESP:T19D12.1

A:Experimental source: strain Bristol N2; clone t19d12

C:Genetics:

A:Gene: CESP:T19D12.1

A:Map position: 2

A:introns: 36/1; 134/2; 180/1; 622/3; 691/2; 754/1; 1111/2; 1174/1; 1271/3; 1322/2; 1681

Query Match

Best Local Similarity 4.9%; Score 210.5; DB 2; Length 1777;

Matches 162; Conservative 100; Mismatches 266; Indels 205; Gaps 38;

Qy 2 TLVTKVAPVAPVSSGPRLLPAPQIV-AVKANNTTIOFPANLQLPQGVILKSNQPL 60
Db 453 TLQSPTPTSTPLVSSSSGSSSTVTTSTPTSTQGVPTSTNQTPST-----SNP- 505
Qy 61 MLVSPQOVTTRAETTSNITSRPAVPANPQTVKICTVPS-----SSQLKKAIVTPVKKL 115
Db 506 -TTPKSTVAPSTGANS---TASPTSTI-TSAPTSQSHSPSSSTMTSTVPT--STP 556
Qy 116 AQTGTIV-VTVPKPSSVQS-VAVPTSVTVT-----PKKPL----- 150
Db 557 APTTTPVTPVAPGQCYCQSNVAVAFELTSGTSDLDLDIONFTANTLFTYSGAPYLLGL 616
Qy 151 ---NVTTLKP-----SLGASSTPS-----NEPNLKAENSAVQ--INTSPM 189
Db 617 TONRTATSLVPPNDTLDLMTYGAEOPTPSGALSALDFTNILARGNAVLSDAFNITP-- 674
Qy 190 LENVKCKNFLMLIKLACSGSPSPMGONVKVLEQLDA-KIBAEETRLKLYELKS 247
Db 675 --NTR-KGYGFFVLVA-----NSDESQVQSDATNLKAGF-NVITVAFS 719
Qy 248 S-----PQPH---LVFLKKSVA-----LRQLLPNS-----QSFIOQ 277
Db 720 SKKFDVLASQPSYNTIYODADKRYATLIGNVLTNTYCNMLAASSTMTSGVYSSTVY 779
Qy 278 CVOQTSDDAVIATCT--TVTTSPVYTTT--VSSQ---SEKIIYSGAT-APRT--VS 326
Db 780 TTQSGSSQAPSTVYPTTGTSAGASTIGSTIOQASTSSVITGTSAPQSSSTANS 839
Qy 327 VQTLNPLAGVAKAGVTLHVGPTAATGTTAGTLQTSKPLVTSVANTVYVSLQ- 385
Db 840 STTTSF-----STTGSTPAPOSSIVASTVTSPTTECTICTVYNTFGTSTTG 890
Qy 386 -----PEKPVSGTAVTTLPLAVTBETSGAICLPSVAPVVSFCMDHICKPVIGTIPVQI 440
Db 891 LTSSSQASTAGSVSTVASTTIPQSSS-----SSPOSP 925
Qy 441 KTAQGPVLSQAPAGIPTGSSSKQLFSLFHVVOQPSGCKNEKQVTTTSHSLTLIQKG--- 497

Db 926 TSQAPQSSSTAAATYVSSS-----QSFSTSPAQSSSTPAQSTVYVQSSSFQ 975
Qy 498 -QKMPV--NTIIPTSQPPASILKQITLPGNKILSLQASPTQKNIKNTVSCFDED 553
Db 976 SPQSTQIGSSTVTPSTQVQASST-----SGPPTTQICPNQOVFKQGVYI---E 1024
Qy 554 DINDVYMAQVNLNEENACILATNSLVLQSCADEPLFLGAL-----QKRIIDIG 607
Db 1025 MLPASTQQNAIMAFEVNVL--NSMEYGLALDNLTHDNRTLVTAIYPTTDYKMYOGYG 1081
Qy 608 KKHDTLNSDAV 620
Db 1082 SANSVDEFKQVI 1094

RESULT 13

A47283

calphotin - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 21-Sep-1993 #sequence_revision 25-Apr-1997 #text_change 21-Jul-2000

C:Accession: A47283

R:Ballinger, D.G.; Xue, N.; Harshman, K.D.

Proc. Natl. Acad. Sci. U.S.A. 90, 1536-1540, 1993

A:Title: A Drosophila photoreceptor cell-specific protein, calphotin, binds calcium a

A:Reference number: A47283; PMID:93165730; PMID:8434015

A:Contents: photoreceptor cells

A:Accession: A47283

A:Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-873 <BAL>

A:Cross-references: GB:U05080; NID:g157071; PIDN:AAA28420.1; PID:g157072

A:Note: sequence extracted from NCBI backbone (NCBIN:124958; NCBI:124959)

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo

Query Match

Best Local Similarity 4.9%; Score 210; DB 2; Length 873;

Matches 118; Conservative 82; Mismatches 239; Indels 110; Gaps 22;

Qy 3 LTVKVAVAPVAPVSSGPRLLPAPQIVAVKAPNNTTIOFPANLQLPQGVILKSNQPL 62
Db 29 VVSPAAVAVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 87
Qy 63 VSPQOVTTRAETTSNITSRPAVPANPQTVKICTVPS-----SSQLKKAIVT-----PVKKLA 116
Db 88 VAP--PVVAPPTP-----AASPVSTPVAVADIPVAVSAPVAPVAPVAPVAPVAPVAPV 140
Qy 117 QIGTT--VVTTPKPSVQSAVAPTSVTVTPCKPLMTVTTKLPSSIGASSPSPNEPNK 174
Db 141 PVATPVPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 195
Qy 175 AENSAVQINLSPMLE-----NVKCKNFLMLIKLACSGSGSPSPMGONVK 221
Db 196 PVLAAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 243
Qy 222 KLVLEQLDAKIEAEETRLKLYELKSSPOPHLVFLKKSVALROLPLNSQSTIOQVQ 281
Db 244 KPLAAEPVVPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 290
Qy 282 TSSDMVATCTTNTVTVSPVTVTVSSQSEKSIIVSGATPRTVSVQTLNPLAGVAKA 341
Db 291 TEPVPAATITLTPER-PALAPVYAESQVAAVNAVATPPRP-APEETIAP----- 339
Qy 342 GVTTLHVGPTAATGTTAGTLQTSKPLVTSVANTVYVSLQPEKPVVSGTAVTSLP 401
Db 340 -----PVVAETPEVAVAAETTPPVVPA-----ASIDAPVATTPVATL- 383
Qy 402 AVTFGETSGAICLPSVKPVVS-----FCMDHICKPVIGTIPVQIKLAQGPV 448
Db 384 AVTDPPVTSASV--PELPIAPSPVPSAVAETRPVDLAPVLPVAPVAPVAPVAPVAPV 441
Qy 449 LSPAGIPTGSSSKQLFSLFHVVOQPSGCKNEKQVTTTSHSLTLIQCGKQKMPVMTIIP 508

Db 442 TPAPASAPVTIALDIPVAVIAPSDAPAE-----APSAAPI---VSTPPTTASVP 492
QY 509 TSOPFPASI 517
Db 493 ETAPPAV 501

RESULT 14

calcium-binding protein calphotin - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C/Accession: A47282
R:Martin, J.H.; Benzer, S.; Rudnicka, M.; Miller, C.A.
Proc. Natl. Acad. Sci. U.S.A. 90, 1531-1535, 1993
A:Title: Calphotin: a *Drosophila* photoreceptor cell calcium-binding protein.
A:Reference number: A47282; MIMD:93165729; PMID:8094559
A:Accession: A47282
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-865 <MAR>
A:Cross-references: GB:L02111; NID:9157031; PIDN:AAA28405.1; PID:9157032
A:Experimental source: photoreceptor cells
A>Note: sequence extracted from NCBI backbone (NCBIN:124955, NCBIPI:124956)
C:Genetics:
A:Gene: FlyBase: Cpn
A:Cross-references: FlyBase: Fggn0010218
C:Superfamily: collagen alpha 1(I) chain, fibrillar collagen carboxyl-terminal homology
C:Keywords: calcium binding

Query Match 4.8%; Score 206.5; DB 2; Length 865;
Best Local Similarity 21.7%; Pred. No. 0.0025;
Matches 122; Conservative 82; Mismatches 242; Indels 115; Gaps 22;

QY 1 GTLVTKV-APYSAP-----PKYSSGRRLAP---QIVAVKANPTTIOFPALQ 45
Db 4 GTTSPVSAIPVAPVTSVSAVAPVQVSPAAPVAPVAPVAPVAPVAPVAPVAPVAPV 63
QY 46 LPPTVLKSNQSGPLMLVSPQVTVRAETTSNITSRPVAPVAPVAPVAPVAPVAPV 104
Db 64 IPAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 115
QY 105 KKAIVTPVKKL-AOIGTTVTTVPKPSVSVAVPTSVT-VTDGKPLNTVTLKPSGLG 162
Db 116 PPAVATPTVQVQVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 175
QY 163 ASSPSNEPNLKAENSAVQVNLSPMLE-----NKKCKNPLMLKILACS 209
Db 176 VAAVPAVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 227
QY 210 GSQSPENQVANKLVEOLLDAKIEAEFTKRLVELKSSQPHLVPLFKKSVALLQLLP 269
Db 228 ----PEVSVATKRLAAAEFVAVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 278
QY 270 NSQSFIOCCVOQTSSDMVATCTTTVTTSVPTVTTVSSQSEKSIIVSGATAPRTVSQT 329
Db 279 VSAS-----TEPPVAAATLTAPET-PALAPVAVESQVAAVAVVATPTPT-APEDET 328
QY 330 LNPAGVGAAGAVVTLHSGVPTAAGTGTAGTGLQTSKRLVTSVANTVTVTSIQPKP 389
Db 329 IAP-----PVVAETPEVAVSAVAETTPVVPVVA-----AESIPAP 364
QY 390 VVSGTAVTLSPAVTFCETSGAALCLPSVKPVVS-----FCMDHICKPVIGT 436
Db 365 VVATTPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 421
QY 437 PVQIKLAPGPFVLSQPAIGIPTGSSSKOLFSLFHVVOQPSGGENKQVTTISHSITLTQKC 496
Db 422 PVPAVVAEEETPEPAPASAPVTLALDIPVAVIAPSDAPAE-----APSAAPI--- 473
QY 497 GQKMPVNTIITPSQFPASI 517
Db 474 -VSTPPTTASVPEVETAPPAV 493

RESULT 15

hypothetical protein K06A9.1a - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C/Accession: T34433
R:Geisler, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of *C. elegans* cosmid K06A9.
A:Reference number: 221525
A:Accession: T34433
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1032 <GRT>
A:Cross-references: EMBL:U08046; PIDN:ACT0869.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 839/1; 849/1; 875/3; 940/2; 966/2;

Query Match 4.8%; Score 206; DB 2; Length 1032;
Best Local Similarity 22.5%; Pred. No. 0.0034;
Matches 141; Conservative 79; Mismatches 246; Indels 160; Gaps 26;

QY 2 TLVTVAVASAPPKYSSGRRLAP---QIVAVKANPTTIOFPALQPLPGTVILKNSG 58
Db 246 SISTSALPI-ASSSASSPSAASSTTPVLLSSSTIQSSSGTFFSVASSPSTV-GSTG 302
QY 59 PLMLVSPQVTVRAETT-SNITSRPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 117
Db 303 AASSSVATVSTIAGTSTSTT---PVGSSSTISSSTPSASSSSGTMTIS-----GS 354
QY 118 IGTVVTVTPKPS-----VQSVAPVTSVTVTPGKPLMTVTLKPSIGLASSPSE 170
Db 355 TGST-VTVVPSSTSTFFASTPIASSSSPGSTVAVAGS-----SSTVGSSTPS- 401
QY 171 PNLKRNAAVQVINSPTMLEVKKCNFLMLILKACSGSPMGQVKKVQQLDA 230
Db 402 -----ASSSSSGTMTNSGSTGTVA 424
QY 231 KIEAEFTKRLVELKSSQPHLVPLFKKSVALLQLPNSQSFIOCCVOQTSSDMVAT 290
Db 425 PVSSSTFG-----SSTP-----IASSSSGSTVTVVSSSSTYGSS 460
QY 291 CTTVTSPTVTVTTVSSQSEKSIIVSGATAPRTVSQTLNP-----LAGPVGARAG 342
Db 461 TPSASSSAGVASTISGSTGATATVPGSSSSVSGSTQASAPSPGTMTSVSGPTGTYT 520
QY 343 VVTLHSGVPT-----AATGCTAGTGLQTSKRLVTSVANTVTVTSIQPKP 391
Db 521 VVPGSSTSPADSSSPNPSSTPASTSTTITISG---SSSIIVTSVGS-----TV 566
QY 392 SGTAVTLSPAVTFCETSGAALCLP-SVKPVVSFCMDHICKPVIG--TPVQIKLAPQPV 448
Db 567 SGTSTISOSTIASSATATGSSSTVSSSSPQPS---SOSPAPNCTGTPSTSSSPSPS 623
QY 449 LSQPAIGPTGSSSKOLFSLFHVVOQPSGGENKQVTTISHSITLT--OKCGQKMPVNTI 506
Db 624 MNPSSSTPTGSSQSTITPEGSTASSPTG-----STGTFVAVTAVTQSTVPSGSS 674
QY 507 I---PTSQFPASITLKOTTLGKNTLSIQASP--TQKRIKENVTSCTREDIDIVDTVM 561
Db 675 LGTOSTNSPSPSSLSPT-SGMSTLTSEPSSTQSSGGAOSTLT-----PSPNPQST 728
QY 562 AGVNLNEENACLIATNSLAVGLTLOS 587
Db 729 SLESSTSGA---TTSSSGAGTMTS 751

Search completed: February 16, 2003, 22:00:16

Thu Feb 20 16:36:13 2003

Job time : 51.8257 secs

us-09-763-909-2.rpr

Page 10


```

Db 108 TTVQPTTTLQSRPGVQPOLVLGSSAQPASLGTATNAVOTGRQRTVPGASTTSTAATF--- 164
Qy 186 SPMLEENKCKKFLMLILKACSGSOSPEMGONKKLYEQLLDIAKEAEETFRKLVEL 245
Db 165 ---TMENVKCKCKSEFLTLILKASSGOSKETANVONLDVONLDGKIEADEFSRLREL 221
Qy 246 KSSPOHVLPELKKSVALLKQLLPNSOSFIOGVQOISSMVAITCTTTTSPVYTVV 305
Db 222 NSSPOYVLPFLKRLPALMLQLTPDSNAFTQSSQOQPP---ASQATTLATNAVLSVV 277
Qy 306 SSSQSEKSIIVSGATAPRVYVOTLNLPLAGPVAKAGVYVLSVGPATAGTTAGTGL 365
Db 278 QRTAGTAAVSATLQPPVLSL-----TQPVGVGKQ 310
Qy 366 QTSKPLVTSVANTVTVSLOPEKPVVSGTAVTSLPAVTFGEFGSGLICLPSKPVVVSFC 425
Db 311 APPTPVITQ-----QPPKP---GALLRPPQVITLQV-----PVALR 344
Qy 426 WDHCIPVIGTPVQIKLQAGPVLSQAGIPITGSSSKQLSLFHVYVQPSGKNEKVTTI 485
Db 345 QPH-NRIMLTTPQOIQUL----- 360
Qy 486 SHSSTLTLOKCGQKTPVNTIIPTSQFPASILKQITLPGNKIL---SLQASPTOKNRK 542
Db 361 -----NQLQPVVVKPTVLPBETKALSTVSAQAAAAQKNNKK 396
Qy 543 ENVTSCFRDEDINDYVTSAGVNLNEBNACILATNSELVGTLIQSCDEPFILGALQKR 602
Db 397 EPGGSGFRDDDDINDVASMAGVNLSESAIILATNSELVGTLTRSCDDPFLLPAPLQRS 456
Qy 603 IIDIGKKHDTLNSDAVNLISOATQERLGLLEKLTALIOHMTYKASENTYILCSDFR 662
Db 457 TLEIGKKGHTDCHPVDVSVSHATQORLQNVKEKISERAOQKNFSTKDDREQASDVR 516
Qy 663 SOLFLEKLDQLEKORKLEEREMILKAASRSNKEDPEQLRLKOKAKELQOLELAQIHR 722
Db 517 AQLKFEQLOQJOKRKDEQERELMRAKRSRQEDPEQLRLKOKAKEMQOQELAQMRQ 576
Qy 723 RQNLATATATGPRKKRPLE-----SGIEGLKDNILASGSSLTATQOLHRPITRIGLR 777
Db 577 RQNLATATATGPRKKRKYDCTGTGSAQSGSPGAAPVGGSGVGTPOFROTITRVNLR 636
Qy 778 DLIFCMEQEREMKYSRALYLAL 800
Db 637 DLIFCLENERETSHLLLYKAF 659

RESULT 2
Q8T9E0 PRELIMINARY: PRT: 851 AA.
AC 08T9E0:
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE SD04735P.
GN TAF110.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Broksstein P., Hong L., Agbayan A., Carlson J.,
RA Champagne M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuncio J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,
RA Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY069807; AAL39952.1;
SQ SEQUENCE 851 AA; 92093 MW; 665B28B9588C984C CRC64;

```

Query Match 17.7%; Score 753.5; DB 5; Length 851;

```

Best Local Similarity 30.5%; Pred. No. 1.6e-34;
Matches 263; Conservative 122; Mismatches 291; Indels 187; Gaps 31;

Qy 22 LPAQIVAVK--APNTTIO--FPANLQPPGTGLIKNSGPMILVSPQOTVTRAEETSN 78
Db 93 LPAQVYVGMQAPSPQOQKQNMPTN---PLSRVYNHMGAVRQSP--SSMTNTYATANSI 148
Qy 79 TSRPAPVAPNQYKICTVPNSS--QLIKRVATVPKKAQIGTVTVTPKPSVQOVA 136
Db 149 I-----VNSVASSGVANSSQPPHLLQNLNAP--QLPQI--TQIOTIPAOOSQ--- 193
Qy 137 VPTSVVTVTPGKPLNTVTLKPSLSGSSTPSPNEPNLKAENSAVQINLSPTMLEANK-K 195
Db 194 -----QQVNVNNSAGTATVASTTA-----ATTQGNTEK 226
Qy 196 CKNFELAMLIKACSGSOSPEMGONKKLYEQLLDIAKEAEETFRKLVELKSSPFLVP 255
Db 227 CKRFLANLIEL--STRPKRVKKNVRLIQLIVANANPEPEFCDRLERLNASPQOLIG 284
Qy 256 FLKSSVALPOL-----LPSQSPITQ-----QCVOQ 281
Db 285 FLKKSIFLLQALYTKELVIEGIKPPQHVLAGLSQQLPKIOAQIRPIGSGQTTIGQ 344
Qy 282 TSSDMVATCTTIVTVTSPPVTTVSSQSEKSIIVSGATAPRVY---VQTLNPLAGPVG 338
Db 345 TQVRMT--TNNALGTPRPTIGHTTISKQPN---IRLTPAPRLVNTGIGITQIP-SLOVP 398
Qy 339 AKAGVYVLSVGPATATAGTTAGTGLQTSKPLVTSVANTVTVSLOPEKPVVSGTAVTL 398
Db 399 GQANIVQIR--GPHQALQATGSGVQIRATIRP-----PNSVPRAN----- 456
Qy 399 SLPAVTFGEFGSGLICLPSKPVVVSFCWDHCIPVIGTPVQIKLQAGPVLSQAGIPITG 458
Db 437 KLTAVKVGQTOIKAI--TPSHLP-----PSLAISGSP-----PPTPLVSLSTLSA 482
Qy 459 SSSK--QLPSLPHVYVQPSGKNEKVTTSHS-----TLT 482
Db 483 STTTLTPSLPPLVHLPEALRAREQMONSLNHSNHFDAKLVEIKAPSLHPMERIMAS 542
Qy 493 IQRCQKQTMPEVNTIIPTSQFPASILKQITLPGNKILSLQ-----ASPTOKNRKE 543
Db 543 LPTIGAKTM-----APPRAINKALIGKKRKDMEMDAKINTSSGGAASANSFQO 593
Qy 544 NVTSCFRDEDINDYVTSAGVNLNEBNACILATNSELVGTLIQSCDEPFILGALQKR 603
Db 594 SSMSSMTGDDDDINDVAMAGVNLAEESQRIIGC--TENIGTQIRSCXDEVFLNPLSQARI 652
Qy 604 LDIGKKHDTLNSDAVNLISOATQERLGLLEKLTALIOHMTYKASENTYILCSDFRS 663
Db 653 RAITSEAGLDEPSODVAVLISHACQERLKNIVERKLAVIAEHRIDVLIKDRYPAPKDVGR 712
Qy 664 QLFLEKLDQLEKORKLEEREMILKAASRSNKEDPEQLRLKOKAKELQOLELAQIHR 723
Db 713 QIFLELELDAEQKRHELEEREMILRAKRSRQEDPEQAKKAKAKEMQRAEMELRQR 772
Qy 724 DANLTATATGPRKKRPLE-----SGIEGLKDNILASGSSLTATQOLHRPITRIGLR 778
Db 773 DANLTATATGPRKKRLKIDETVSSGSSGGVLSGSSGAPPTL-----RPRIKRVNLR 828
Qy 779 LIFCMEQEREMKYSRALYLAL 801
Db 829 MIFYMEQEREFRCSSMLFKTYLK 851

RESULT 3
Q61707 PRELIMINARY: PRT: 549 AA.
AC 061707:
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, last sequence update)
DE 01-MAR-2002 (TREMBlrel. 20, last annotation update)
DE Hypothetical 60.4 kDa protein.
GN R119.6.

```


Db	2126	PAKAOPOTASAPRCPOTOPQSPAPREVOOTQOTVOTSSHVPEAKOPTHA-----	2179
Qy	361	GTGLQTSKPLVYTSVANTVTTVLQREPKRVSTAVTSLPVTGEGTGAATCLPSKVP	420
Db	2180	-----QSSKPDVAOAS-----QOSNVOGOSPVRVOSPQOTIRPSTPSQLSPGQOS	2226
Qy	421	VVSFCMHCKRVPYGVQIKLQAPRPVLSORAGITSSSSKQSLSLFHVHQOPSGGNEK	480
Db	2227	QVO---TTTSOPi---PIQPHSTLQIPSGQOPSPQOVQOSSPTQLS-----SGOTLN	2272
Qy	481	QVTTISHS-STLTQKCGQKTMVNTPIPTSGF-PRASILKQITLPGNKIISLQSPYOK	538
Db	2273	QVSVSSSSRQLOLQO-----PQPVYAVPQLOOVYLSOI--QSOVVAOIA--QO	2321
Qy	539	NRIKENY-----TSCFRDEDDINDYTSNAGVNLNEENACILATNSELVTLQIOCK	589
Db	2322	SGVPQQIKLQLPIDQOSSAVQFHQJQNVYVQAASVQQL-----GRVQOLR	2369
Qy	590	DEPLFTGALQKR---ILDIGKKNHITELNSDAVNLISGATQERLGLLEKLTALTAQH-	644
Db	2370	DOQ-----OKKKOQOIEIKREHTLQASNQ-----SEIIOKQ---VVMKNHVAIEHLK	2413
Qy	645	---RMTYTKASEN---YILSCDTSQKLEKLDQLEKQKXDELEEKMLKAAKSSNKED	699
Db	2414	QKKSMTPAEREEENOMIVCQOVAKYI--LDKIDKEKO-----AAKKRRREES	2459
Qy	700	PEQRLRKQAKELQOL	715
Db	2460	VEQKRSKQNTKLSAL	2475

RESULT 5	Q08TE50	PRELIMINARY;	PRT; 1322 AA.
ID	Q08TE50		
AC	Q08TE50;		
DT	01-JUN-2002 (TREMBLrel. 21, Created)		
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	EMSY protein.		
GN	C110RF30.		
OS	Homo sapiens (human).		
OC	Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_Taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Hughes-Davies L.;		
RT	"EMSY is amplified in breast cancer and displays a BRCA2 dependent DNA		
RT	damage response.";		
RL	submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.		
DR	EMBL; AJ30203; CAD22828.1;		
SR	SEQUENCE 1322 AA; 141467 MW; 7F8C95EBBA0FC9F0 CRC64;		
Query Match 5.9%; Score 250.5; DB 4; Length 1322;			
Best Local Similarity 20.0%; Pred. No. 7.4e-06;			
Matches 207; Conservative 157; Mismatches 370; Indels 301; Gaps			
QY	5 TKVAVSAPRKVSQSGRLPAQIVAVKAKAPNTTIOFPAN--LQIPPG--TVLIKNSGFLM 61		
Db	149 TTSTPISTP-----VPSGSIATVKSPPRAS--PASNVLPLSGSTVYKVSQSCDE 197		
Y	62 LVSPQQTVTAEFTTSTNITSKPAVPANPQV---KICVTPNS--SOLIKKVA----- 108		
Db	198 DEKRK---RRRINSSSSSVLKEVPKAVVSKTITVPVSGSPKKNIMQSTANSLPP 254		
QY	109 -VTPVK-----KLAIGIT-----VTTVPKPSVQSV-----AVPTSVTV 144		
Db	255 HMSVKKITTEPKPSTQTTNTTQKVIITVTSPSTFEVPNILSKSHYAAVTKLVPSTVIA 314		
QY	145 TPQGLMTVTTLKPSL-----GASSTPNSMPPN-----LKAENSA 180		
Db	315 TIOKRP--PVVITASQSSLSVNSSGSSSTSPSPIPNTVAATVAVSTSPVMSITVAQGVST 373		

QY	181	VOJNLSPEMLBNVCKCNFLMAMLIKACSSOSEMGOANKKLVEJLDBLAKIEA -BEJFR	239
Db	374	SAIKMASTRLEPSKSLVSAPIQOLI-----AOEKKHQOSPK--QOLUYOQOQTQOOVAQ	425
QY	240	KLYVELKSSPOPHVLPFLKRSVVALROLLEPNSOEFIOOCVOQTSMDMVAITCTT-----TV	295
Db	426	PSPVSHOQOPOQSPPLPGIKRTIQIKQ-----ESGVIIITQYQVPSKILPKPVTATLPS	480
QY	296	TTSPPVTTTIVASSOSEKSIIVSGAFAPRVYVQTLNPLACPVCAKKGVYT-----	345
Db	481	SNSPIWVSSNGAIMTKLVTTPGTQATYTRPVSPIGRMATGCAATVYKTTSGSII	540
QY	346	-----LHSV-----PLAATGGTTAGTGLQTSKPLVTSVANTV-----TTSVLQPEKPV	391
Db	541	TVPKSLATLOGKIISSNIYSGTTTKITTTIPMKSFPVIVYQKTTGKTIGLOPKKNV	600
QY	392	S-----GTAVTLSPAVTGETSGAALICLPSVKPVVSCFMDHCKPV-IGTVPQIK--	441
Db	601	TTLNAGGEKTIQTV-----TGAKPAITLATRPITKMI-----VTPKGIQSTVCPAAKI	651
QY	442	-----LAPQPVLSQAPGIFITGSSNQPLSFHVQOPS-----	475
Db	652	IPYIIVGQOGKTOVLIKRPVYTFQATV--SEQTRQVLT-ETLQOARVVAEAGNSSIQ	707
QY	476	-GGENKOVTTTISHSSTLTIOKCGQKTPVNTIIPT-----SOEPASILKROIT	523
Db	708	EKGREPONYTDDSSSSSTRESSOSSODSQPVYHVIVLASRQDMSEHEIMETSPTIIVDVSS	767
QY	524	PGNKILSLQMS--PTQKNRIKENYTSCEFRDEDDINDVTSMA-GVNLINE-----ENACIL	574
Db	768	ESOGATSTIKALLETLOOTYKWKLES--KPROPTIDLSQNAVIOQTOEKRHSPESSPIA	825
QY	575	ATNSEIVGTLIGQSKDEPFLFIGALOKRIIIDKHDIFELNSDAN-----LISAQO-	628
Db	826	VVESELVAEYI-----TTERIDEGTEVAFPLLVHSRSOP	859
QY	629	-----ERLGLLEKLTIAIOHRMTTYKA-----SENYILCSDRSOLKFELEKD	672
Db	860	QOPSQOPRTLLQH--VAOSQTAQTSVVVVKSIAPASPGALITHIMOALSSHNAFTKHS	916
QY	673	OLEQORKDLEREML-----LKAKSSNKEDP-EQJRL-----KOKA	709
Db	917	ELGTEGEVEEMDPLDBQIGLFYRSALTOSQSAKOQKLSOPLEQTOLOVKTLOCFOTQO	976
QY	710	KELQOLELAQIOHNDAMITTAATIGPKRKRPLSGIEGLKDNLTLAGSTSLATKOLHRP	769
Db	977	KQTHLQANDQOLQHLPPM-POLSTRHKLTLPLOEOAOQRPDY-----QHTQHP	1022
QY	770	RITRICALDLIFCMEQREMKYSBALYLALLKZPLSHSIHILAIYCORRHKALLMC--P	826
Db	1025	MYAK-----DROLP-----TLMAQPPQTVVQVILAVKTTQOLPKIQQAPNPQ	1066
QY	827	EISISGKZHOHERAL	841
Db	1066	KIYVOPQTPQSOXSL	1080

RESULT 6	
P70670	
ID	P70670
AC	P70670;
DT	01-FEB-1997 (TREMBLrel. 02, Created)
DT	01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE	MASCERT polypeptide-associated complex alpha polypeptide (Alpha-NAC, muscle-specific form GP220).
DE	NACA.
GN	
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX	NCBI_Taxid=10090;
NN	[1]


```

OY 512 FPPASI-----LKQITLPGNKILSLASPTOKRIKENVTSCFDEEDINDVTSMAGV 564
DB 949 NPPCFHENGTNTTATTYYAN--LGGHPQPTOVQFV-----CDKQETASLVTSAVG- 999
OY 565 NUNEACILAINSELVGLTQSCNDP 592
DB 1000 ----QON-----GNVVRVCSNMP 1013

RESULT 8
OQFMHO PRELIMINARY; PRT; 689 AA.
AC 09FMO:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE GblAAT24960.1.
OC Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX MEDLINE=98162728; PubMed=9501997;
RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.
RT Sequence features of the regions of 1,191,918 bp covered by seventeen
RT physically assigned P1 clones.";
RL DNA Res. 4:401-414(1997);
DR EMBL; AB008267; BAB08277.1; -
SQ SEQUENCE 689 AA; 75680 MW; 4A028C47111AFA3E CRC64;

Query Match 5.5%; Score 234; DB 10; Length 689;
Best Local Similarity 19.7%; Pred. No. 2,7e-05;
Matches 143; Conservative 125; Mismatches 258; Indels 200; Gaps 29;

OY 167 PSENPFLKAKNSAIVOI---NLSPWLENVKCKCNFLAMLIKACSGSQSPKGNQVKKL 223
DB 64 PVNPMNVNFINRNPKOVFPFALLPILMNQDK-----DRALQLRTL 104
OY 224 VQOLIDAKLEAEFTKRLVELKSSQPHLVPFKKSYVALQQLPNSQSFQOCYQORS 283
DB 105 YARKLKNELPKGFTKRM-----KDIVG-DQMLKMAVSKIQ-QVNYNQ 145
OY 284 SDVIATCTTCTTSPVVTTVSSQSEKSI-----VSGATAPRTVYQTLN----- 331
DB 146 KGIGIQAPSTELNN-----QKQSDPRAVHLNQLPSSAGTIGSSVPVQGLTKRHPQ 197
OY 332 ----PLAGPVGAKAGVTVLHSV-GPTAATGTTAGTGLLOTSKPLVTSVANTVTVTSIQ 385
DB 198 QMOHPPSPSPMYTSG--SFHSFGPNTNAGSTLRLPHLDHSH---MRVHANNQPMGSG 252
OY 386 PEKPVVSGAVVLSLPAVTFGFTSGAICLPVKRVVSCWCHICKPVYIGTVQIKLAOP 445
DB 253 LGGPOSTTNM-WTWP--KFERPSS-----VNDSTRV---OG 283
OY 446 GFV--LSOPAGIPTGSSSQQLSLFFHVVOQPSGNEKQVTTISHSSTLTIQKCGOKTMP 503
DB 284 GATSHFQUNSSSLPLMSAPQGGSSVSHVKQESVDQSEFKNNAASMTSNEDLEKSSR-MVL 342
OY 504 NTIIPTSQFPPASILKQIT-----LPGNK-----ILS 530
DB 343 ST--DNNAAPASSVSPMTQLDASTMNSRGLGTSOGAGANARPPKPPVQKKPLET 400
OY 531 LQASTQKNRIKENTSCPRDE--DDINDVYSMAGVNLNEENACILIA-----TNSELVG 582
DB 401 LGSSPPPSK-KQKVAQSMQDSIDQLNDVYAVSGVNLREEEQLFGSAKEDGRVSEASR 459
OY 583 TLISGCKDEPFLFIALGKRIIDICKKHIDITELNSDVAVLISQAIQERLRGLLEKLTALA 642

```

```

DB 460 RVHFEERELLQKNLQKRIAEIMAKGIQJISNDVERCLSCVEERNGLSHITIRIS 519
OY 643 QHRMTTYKASENVIILSDPRSOLKPLEKLDLEKQKRDLEERMLKAASRKNKEDPQ 702
DB 520 KQVDAEKSHHRTITSIDRLQINEMNQVKKEEMKQKQAEKLT---KKPSEKNKEDDK 576
OY 703 LRLQKAKELQQLLELQIQRDANLTATAIG-----PKRRPLESGIEGLK 749
DB 577 MRTTA-----ANVAAARAAGGDDAFELKQWLMAEAKQSVSEAGKDGND 619
OY 750 DNLLASGTSLSL-----TATKQLR-----PRTR-ICLRDLIFCQDEEREK 790
DB 620 KTTSGGKKSRODQGRFSGTGRVKGKNGSSIQPKVVRTISVDVAVALEREPOKS 679
OY 791 YSRALY 796
DB 680 KSTLMY 685

RESULT 9
OQ29071 PRELIMINARY; PRT; 528 AA.
AC 029071;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Gastric mucin (Fragment).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_Taxid=9823;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE-GASTRIC MUCOSA;
RX MEDLINE=94102478; PubMed=7755593;
RA Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
RA Lamont J.T.;
RT "Isolation and characterization of cDNA clones encoding pig gastric
RT mucin.";
RL Biochem. J. 308:89-96(1995).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN-YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE-GASTRIC MUCOSA;
RX MEDLINE=94102478; PubMed=7506218;
RA Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
RA Lamont J.T.;
RT "Pig gastric mucin: isolation and characterization of a cDNA clone
RT with a novel tandem repeat.";
RL Gastroenterology 106:200-200(1994).
DR EMBL; U10281; AAC48526.1; -
FT NON-TER 1
FT NON-TER 528
SQ SEQUENCE 528 AA; 49907 MW; 0BF0F6879203B2EA CRC64;

Query Match 5.5%; Score 233.5; DB 6; Length 528;
Best Local Similarity 22.3%; Pred. No. 2,1e-05;
Matches 126; Conservative 93; Mismatches 196; Indels 149; Gaps 26;

OY 9 PVSAPKVSQGRPLPAPQVAVKAPNTTIIQFPAHQLPQGVTLKNSGFLMVSPOOT 68
DB 1 PLSVOPSSSSSS--PRTSTTSVSSSSSVPIPTTSVQP-----SSSGS---APTTS 48
OY 69 VTRAEITSNITSPAPVAPNPOTVKKICTVYVNSSSQLIKKVAVPVKLQAOIGTVTVTPK 128
DB 49 ATSVQTS--SSSPIS--TISVQTSSSS-----VPTTSTSV-Q 85
OY 129 PSSVQAVAVTISVVT-----VTGKPLNTVTTTLKPSLSGASSTPSNEPLKAENSAVOI 183
DB 86 PSS--SSSAPTRATRSVQSSSSSAPISSTTSVQSSSGSVYTTATSVQSSSSSA--- 140
OY 184 NLSPTMLENVKCKNFKLMLIKACSGSQSPKGNQVKKLVQQLDAIEAEFTKRLTY 243

```


Db 141 -----TSATSVQSSSSSPPISTV-----SVQSSSSSAPTTSATSV 180
 QY 244 ELKSSPQPHVLPFLKRSVVALRQLLPNSQSFIOQCVOQTSDDVIATCTT---TTSVPV 300
 Db 181 QPSSSSSPPI-----SSTVS-----VOTSSSSVPTTSTTSVQSSSSSV 220
 QY 301 VTTVSSSSQSEKSIIVSGATAPRTVSQTLNPLAGV-----GAKAGVYTLHSVG 350
 Db 221 PTTSATSVSRSSSS---SSTPIPTSTVSQSSSSSAPTTSATSVQSSSSSPPISTTSVQ 277
 QY 351 PTATGCTTACGTGLOTS-----KPLVTSVANTVTVTSIQPEKPVVSGTAVT 397
 Db 278 PSSSSAPPTTSATSVQSSSSSPPISTTSVQSSSSSPPTSTTSVQSS---SGSAPT 334
 QY 398 LSLPAVTFGCTSGAICLPVSVKPVVSCMDHICKPVIGTPIQIKLAOPGVLSQPAIGPT 457
 Db 335 TSATSVQSSSSS-----PPISTTS-----VQSSSSSSPPTSTTSVQSS 375
 QY 458 GSSSKQLFSLFHVYQVQSGNEKQVTTISHSSTLTIOKCGQKTMPTVTIIPTSQFPASI 517
 Db 376 SSGSAPPTTSATSV---QPSSS-----SSVPTTSATSVSRSSSSSPPIPT---TTSVQSS- 424
 QY 518 LKQITLPNGKILSLQASPTQKNRI 541
 Db 425 ---SSVPTTSATSVQSSSSSPPI 446

RESULT 10

061191
 ID 061191 PRELIMINARY; PRT; 2045 AA.

AC 061191;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Transcription factor C1 (HCF).
 GN HCF1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclerognathia; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RA Kristie T.M., Dasher R.;
 RT "CDNAs encoding the mouse homolog of the human transcription factor C1 (HCF).";
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: u53925; AAB0163.1; .
 DR MGD: MG1:105942; HCF1.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR001798; Kelch.
 DR Pfam: PF00041; fn3; 2.
 DR Pfam: PF01344; Kelch; 5.
 DR SMART: SM0060; FN3; 1.
 SQ SEQUENCE 2045 AA; 210537 MW; B89CC2FDA35B969F CRC64;

Query Match

Best Local Similarity 5.5%; Score 233; DB 11; Length 2045;
 Matches 158; Conservative 85; Mismatches 253; Indels 188; Gaps 36;

QY 7 VAPVSAAPKVVSSGPRL---PAPQIVAVKAPRTTITQFPANQLDPGGVLIKNSGPMILV 63
 Db 486 VLKVTGQATVGTPLVYMRASQ---ACKAPVTYV-SLPASVYM---VVPQSQAGCTVIGS 539
 QY 64 SPOOT-----VTRAEITNSI--TSRP---AVPAN-----PQYKICTVP 97
 Db 540 NPQMSGAALAAALAAATOKIIPSSAPTAWSVPACTTIVKTVAVTPGTTTLPATYKVASSP 599
 QY 98 -----NSSSOLIKKVAV---TPVKKLAQIGTVTVTPKPSVSVASVAVTPVTPGKPL 150
 Db 600 VMSNPATRLKTAALAAQVGSVSSAAMTSTRPIITTVKSGIV-TVAQQAQVYTVTVGCVT 658
 QY 151 NYVTTLK-PSL-GASSTPSNEPML-----KAENSAVAQINLSPTMLENVKCKKFL- 200

Db 659 KTIILVKSPIISVPGSALISMLGKVMYVQTKPVQTSATVQASTGPTQIQTGFLPA 718
 QY 201 AMLIKIACSSQSP-----EMGQNKVLYEOLDLAKIAEEFTRLVLYELKSSPPH 252
 Db 719 GTILKLVTSADGKPTTITTTTQASGAGTKPTLGI-----SSVSPS 759
 QY 253 LVPLFKSVVALRQLLPNSQSFIOQCVOQTSDDVIATCTT---SPVVT 302
 Db 760 TT---KRGTTTIKTIPMSALITIOAGATGVTSSEKISPTIITTKVTSSTGAPAKIIT 816
 QY 303 -----TVSSSQSEKSIIVSGA---TAPRT-----VSVQTLNPLAGPVGAK- 340
 Db 817 AVPKIATGHGQGGVQYVLKAGAPQGTILRTVPMGQVRLVPTVSAVKRAVATVLLVYGK 876
 QY 341 -AGVTLHSVPTATGTTGTLGLOTSKPLVSA--NVTVTSIQPEKPVVSGTAVT 397
 Db 877 TTGVTTLGTVTGIVST--SLAGAGHSTASLAPITTLGTTATLSQ---VINPAIT 930
 QY 398 LSLPAVTFGCTSGAICLPVSVKPVVSCMDHICKPVIGTPIQIKL-AQPGVLSQPA-GI 455
 Db 931 VSAQOTITLTAAGGLTPTITMQPV-----SQPTQVTLITAPSGVEAQPVADL 977
 QY 456 FTGSSSKQLFSLFHVYQVQSGNEKQVTTISHSSTLTIOKCGQKTMPTVTIIPTSQFP 515
 Db 978 PVS-----ILASP-----TTEQPTATVITADSGQGDVQGTVTLVCSNPPC 1018
 QY 516 SI-----LKQITLPNGKILSLQASPTQKNRIENYTSCTREDIDINDYTSAGVNLNE 568
 Db 1019 ETHETGTTNATVTVVAN--LGHPQPTQVQV-----CDROETASLVTSAVG---Q 1066
 QY 569 ENACILATNSLVGLIOSKDEP 592
 Db 1067 QN-----GNVAVCSNPP 1079

RESULT 11

08TDH7
 ID 08TDH7 PRELIMINARY; PRT; 1029 AA.

AC 08TDH7;
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Intestinal membrane mucin MUC17 (Fragment).
 GN MUC17.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-21845432; PubMed-11855812;
 RA Gum J.R., Jr., Crawley S.C., Hicks J.W., Szymkowski D.E., Kim Y.S.;
 RT "MUC17, a novel membrane-anchored mucin.";
 RL Biochem. Biophys. Res. Commun. 291:466-475(2002).
 DR EMBL: AF430017; AAL89737.1; .
 SQ SEQUENCE 1029 AA; 108320 MW; 5E1AD19E4BF948E3 CRC64;

Query Match

Best Local Similarity 5.4%; Score 231.5; DB 4; Length 1029;
 Matches 178; Conservative 113; Mismatches 362; Indels 233; Gaps 35;

QY 4 VTKVAPVSAAPKVVSSGPRLPAPQIVAVKAPRTTITQFPANQLDPGGVLIKNSGPMILV 63
 Db 137 VDRSTPVTTSTQSNSTP--TPPEVITLPMSTPSEVSTPLTI-MPVSTTSVTSISAGTAST 193
 QY 64 SPOOTVTRAEITNSITSRPAPVAPNDQYKICTVNSSSOLIKKVAVPVKKLAQIGTVV 123
 Db 194 LPVDTSTPVITSTVQSSSPVTPPEG-TIMPITV-FSBSGTPLTTPMVSITTRVTSSEGSILS 251
 QY 124 TVVPRSSVSVASVAVTSTVTVTPGKPLNTVT-----TLKPSLSGASSTPSNEPNIKAENS 178

```

Db 252 T-----PSVVTSTPVTTSTAISSTATLDSTMSVSMPEISLTITLIVSTPTPTPES 307
Qy 179 AAVOINLSTPTMLENKKCNFLAMLIKACSGSQSPKMGONAKKVLQDLAKIEBEFT 238
Db 308 TPTST-----PSV-----YTSMSMTTASEGSSSP-----TTLTGT 337
Qy 239 RKLVELKSSPQPHVFLPKKSVALLRLLPNOSQSIQOCVOOTSSDMVATCTTIVTS 298
Db 338 TTMPTSTTERSTLTTLTVL-----ISPISWSPSASTLSTPPGCDTSTPLASTKAGSFTIP 394
Qy 299 PVVTTTSSVSSQSEKSIIVSGATAPRTVSQVTLNPLAGPVGAKAGVTLHVSVPAAATGCT 358
Db 395 AEVTTIRISITSERS-----TPLTTLVST-----TL-----PTSPGAS 429
Qy 359 TAGTGLLOTSKPLVTSVANTVTVTSLOPEKPVVSGAVTLSLPAVTFGTSCAICLPV 418
Db 430 IASTPPLDSTSTTPS-----TDTASTPTIPV-----ATTISVSVITEGSPGTITIPST 480
Qy 419 KPVSPSCMD-----HICKPVI-----GPPVQIKLA-----443
Db 481 -FVTSSTADVFAITTCAGVSTPVITSTELMPTSTSSSTTSTSTKEFTTPAATTAAPLT 539
Qy 444 -----QPCPVLSQAPAGIPTGSSSKQLFSLPHVVOQ-----PSCGNEKQVTTISHST 490
Db 540 YVTMSTAPSTPRTSRGCTTSASTLSATSTPHSTSTVTRPVTPSESSSRPSTITSHITIP 599
Qy 491 LTIQCGCKTMVNTIIPISOPPPA-SILKQITLPGNKILSL-----QASPTQKNRIKEN 544
Db 600 PTPPPASSTPTSTASSTTVNPEAVTMTTRKPTSTRTSTPTPTPTPTPTPTPTPTPTPTPT 658
Qy 545 VTS-----CFRDEDDINDVTSMAVNLNENACIATNSELVGTLIQSCKDEPPLF 595
Db 659 PTSTPTPTPTTCFGD-----GCQNTASCKCKGWTGDKCQC---PMLT 700
Qy 596 IGAQKRI---LDIGKADIT-----ELNSDAVNLIQATQORLRLGLEKTLA 640
Db 701 YELCEEVVSSIDIGPETISAMOMELTVTVSVKFTTELKNISSQGEQKQFTQOMNI 760
Qy 641 IAOHKRTTYKASENTILCDTRSOL-KFLEKIDOLEKOKKDLERKML---LKAASRSN 696
Db 761 V-----VSGIPEYVGVITKRLGVSVEHDVLLRTYTPYKIVLDMNATEVVEKITE 813
Qy 697 KEDPEQLRLKO-----KAKELOLELAO-----IOHRDANL 727
Db 814 KTTTQOIMINDCSDMCMCTNTGTQVQNTTVQYDPEEDCKMAKRYGDFVVEYEDQR 872
Qy 728 TATTAAGPRKKRRLBSGLEGKMD-----NLAAGTSSLTATKQLH 767
Db 873 -----PYCISPCEBGFVSFKKCNLKGCOMSLSGPQCLCTTETH 911

```

RESULT 12

```

ID Q9QWH2 PRELIMINARY: PRT: 2045 AA.
AC Q9QWH2:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE C1 transcription factor.
GN HCFCL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN NCBL_TaxID=10090;
RP SEQUENCE FROM N.A.
RA Kristie T.M.;
RT "CDNAs encoding the mouse homolog of the human transcription factor C1
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U08921; AAD09225.1; -.
DR MGD; MGI:105942; Hcfcl.
DR InterPro; IPR003961; FN_III.

```

```

DR InterPro: IPR001798; Kelch.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF01344; Kelch; 5.
DR SMART; SM00060; FN3; 1.
SQ SEQUENCE 2045 AA; 210546 MW; 7AD38FCD78EABA9C CRC64;

Query Match 5.4%; Score 231; DB 11; Length 2045;
Best Local Similarity 23.1%; Pred. No. 0.00016;
Matches 158; Conservative 84; Mismatches 254; Indels 188; Gaps 36;

Qy 7 VAPVSPKPVSSGPRLL---PAPQIVAVKAPNTTIOFPAALQAPGQVLKSNKSGPLMLV 63
Db 486 VLKATVGPQATGTPPLVTKMRASQ--AGKAPVYT-SLPASVRA---VPPQSQGTVIGS 539
Qy 64 SPQQT-----VTRAETTSNI--TSRP---AVPAN-----PQTKICTVP 97
Db 540 NPQMSGMAALAAATAAQKIPSSAPLAMSVPAGTTIVKTVAVTPGTTLPLAVTKVASSP 599
Qy 98 -----NSSOLIKRVAV---TPYKKLAQIGTVVTPPKSSQSVAVPVSVTVTGKPL 150
Db 600 VMSNPAATRLKTAATAAQVGTSSAANTSTRTITTVHKSQTV--TVAQQAQVTVTVVGAVT 658
Qy 151 NPVTTLK-PSL-GASTPSNEPDL-----KAENSAVQINLSPIMLENVKCKNFL-- 200
Db 659 KTTTLVKSPTISVPGSALISLNKAKVMSVQTKPVQTSATVGAOSTGPVYQIHTKGPLRA 718
Qy 201 AMLIKACSGOSP-----EMGQNVKLVQDLDAKIEBEFTRLKLYVELKSSPOPH 252
Db 719 GTILKLVTSADCKPPTIITTTQASGAGTKPTILGI-----SSVSPS 759
Qy 253 LVFLKKSVALRQLPNSGFIQOCVOOTSSDMVATCTTIVT-----SPVVT 302
Db 760 TT---KPGTTTITKTIPMSAIIITQAGATGVTSBPGIKSPITITITVMTSGTGAAPAKIIT 816
Qy 303 -----TVSSSQSEKSIIVSGA-----TAPRT-----VSQVTLNPLAGPAGAK 340
Db 817 AVPKITNGGQGVTVQVYLKAGARGQGTILRTYPMKGVRLVTPVYSAKPAVTTLVKGG 876
Qy 341 -AGVTLHSVPPTAATGTTAGTGLLOTSKPLVTSVA--NTVTVSLOPEKPVSGTAAT 397
Db 877 TIGVTTILGVTVGVST--SLAGAGASTSASLATPITTLCTIATLSQ---VINPTALT 930
Qy 398 LSLPNTFPGTSGAALCLPSVSRPVSEFCDHICKPVIQIPVQKL--NQGCPVLSQA--GI 455
Db 931 VSAAGQTLTAAGGLTPTPTTMOV-----SQPTQVTLTAASGVAQAQVHDL 977
Qy 456 PTGSSSKQLFSLFHVVOQPSGNEKQVTTISHSTLTIOCKGKTMVNTIIPTSQPPA 515
Db 978 PVS-----LLASP-----TTEQPAATVITINDSGGDVQPGTVTLVCSNPPC 1018
Qy 516 ST-----LKOITLPGNKILSLQASPTQKNRIKENVTSQFRDEDDINDVTSMAVNLNE 568
Db 1019 ETHETGTTNATTTVVAN--LGCHPQPTQOVF-----CDROETAASLVTSAVG---Q 1066
Qy 569 ENACIATNSELVGTLIQSCKDEP 592
Db 1067 QN-----GNVAVCSNPP 1079

```

RESULT 13

```

ID 076602 PRELIMINARY: PRT: 1275 AA.
AC 076602:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Hypothetical 122.9 kDa protein.
GN H02F09.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN NCBL_TaxID=6239;

```


RA Cadieu E., Dreano S., Lelaure V., Mottier S., Gallibert F.:
 RT "Sequencing the distal X chromosome of Drosophila melanogaster."
 RN Submitted (JUL-1998) to the EMBL/genbank/DBD databases.
 RP SEQUENCE FROM N.A.
 RA Benos P.:
 RN Submitted (APR-1999) to the EMBL/genbank/DBD databases.
 DR EMBL; AB003421; AAF4564.1; -
 DR EMBL; AL031028; CAAL9845.2; -
 DR Flybase; FBgn0025390; EG:5667.1.
 DR Interpro: IPR002557; Chitin_bind_Pera.
 DR Interpro: IPR002365; P_rich_extensn.
 DR Pfam: PF01607; CBK_14; 2.
 DR PRINTS; PRO1217; PRICHEXTENS.
 DR SMART; SMO0494; ChIBD2; 2.
 SQ SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;

Query Match 5.3%; Score 228; DB 5; Length 1795;
 Best Local Similarity 18.9%; Pred. No. 0.0002;
 Matches 172; Conservative 121; Mismatches 367; Indels 250; Gaps 31;

QY 9 PVSAPPKVSRRPAPQIVAVAKPNTTIOFPANQL-----PPGYLIKNSNG 58
 DB 700 PVSSTGKPTTPK-PSTRPTPTTQITTTPLASSFTSTQPTTTPQPTT 758
 QY 59 PLMLVSPQOTVRAETTSNITSRAVPANQVKTCTPNSSQLIKVAVPVKLAQI 118
 DB 759 TLTLYTPKSTTTTTEKPTSSPK-PTTTOKTSTANTT-----KVAITQKETTPT 812
 QY 119 GTT-----VTVVPRPS-----VQSAVPSTVTVRGPPLN 152
 DB 813 QSTSTTIFRKTNNPEPTSTEKPTSTTPKSTTPKSTTSTVASTTEKTTISSPKPTE 872
 QY 153 VTLKRSSLGASTPSENEPLKAENSAVOINLPTMLENVKCKNFMLIKLACSGSQ 212
 DB 873 KSTENPTNSVKTSAUTSSQRA-----TSTSEPTKTQITTTTPKPTLKNS 922
 QY 213 SEPENQNVKKLYEQILDAR--IEAEFTKLYELKSSPOHLPVPLKKSVALROLNP 270
 DB 923 TQETATSTQKSVTITTKATSESPLLTSTEEPNTTPK-----LRTTTP 970
 QY 271 SOSFIOQCVQOTSSDMVIACTTVTSPVTTVSSSOSEKSIISGA--TPAR--TVS 326
 DB 971 TTS-----VTAFTRTTTTSSSESTSTOKPKSTPTSTRTTPKATTVY 1017
 QY 327 VQTLNPLAGPVGAKAGVTLHSVG-----PTAATGTTA-GTGLLOTSKPL 371
 DB 1018 VSTQNPPT--TTSKSTVTITTPNPSPSTQRPPTTTPROPTSIASSTISGTRIPPTNP 1075
 QY 372 VTSVANTVTVLSQPEKP-----VSGTAATLSLPAYTFCGTSAAICLPSVKFVVSF 424
 DB 1076 QNSTSTDLTITVTRPCPDPDSTSDKNTNTACTQELQOVNLE-----LQSPKQEQF 1128
 QY 425 CWDHICKPVIGT-----PVQIKLAQPGVLSOPACIPITGSSSKOFLSF 468
 DB 1129 TTRTRHTALTGSRNTLIGCEVDPYMDAPSSAFAESQATKAPMSTLAAHLLOKLF 1188
 QY 469 HVV-----QPSGNGEKQVTT-----ISHSSTLTIOKCG 497
 DB 1189 HTTSTTPPREHAPTORPSSQSSSQRSGVTAQMARHNLATSKPFIHSLSLTIOQLA 1248
 QY 498 -----QKTMPTNTII-----PTSOFPASILKQITLPGNKILS-----L 531
 DB 1249 STQKSIPTKTLVNTHTKREPDESLEYDSEYDDEDEVDLDTQPRAMSTTAAVL 1308
 QY 532 QASPTOKNKIKENVTSCFDEDDINDVTMAGVN-----LINEACILA 575
 DB 1309 PAVPTTTRERQKTSSTSPKATSTTQPIETTTGDELDSSSSDYVDAN----- 1363
 QY 576 TNSLVGLTIQSCDEPFLIGALOKRIIDI-----GKKHDTIELNSDAVNLTISQTOER 630
 DB 1364 ---DISSGVNLSLARKNLLSLKQRLQIERTEAKKPATSTSTTADPKTSSSTSP- 1419

QY 631 LRGLLEKLTALIAOHRMTTYASENYILCSPTRSQLFLEKL--DQLEKOKDLEERBML 688
 DB 1420 -----ASTSESTSPVSTYARSKTASHKHLCEALSRKS-----L 1456
 QY 689 KAKKSRNKEDEPQOLR---LKQAKELQOLELAIQIQRDNLNLTATNIGRRKRPLESGI 745
 DB 1457 TPQSAEYEDDDYMEDEPVGSSDAEKKHGVTLISEKQAAATAKNHITAPSAQPLQAM 1516
 QY 746 EGLKDNLLAS 755
 DB 1517 L-----NILAT 1522

RESULT 15
 ID 015052 PRELIMINARY; PRT; 1246 AA.
 AC 015052;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, last sequence update)
 DE 01-DEC-2001 (TREMBlrel. 19, last annotation update)
 GN KIAA0344 protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE=97349984; PubMed=9205841;
 RA Nagase T., Ishikawa K., Nakajima D., Ohita M., Seki N., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. VII.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro."
 RL DNA Res. 4:141-150(1997).
 DR EMBL; AB002342; BAA20802.1; -
 SQ SEQUENCE 1246 AA; 129137 MW; 40CEB1209A245CB5 CRC64;

Query Match 5.3%; Score 224.5; DB 4; Length 1246;
 Best Local Similarity 19.4%; Pred. No. 0.0002;
 Matches 184; Conservative 133; Mismatches 337; Indels 295; Gaps 40;

QY 9 PVSAPPKVS-SGPRLP-----APQIVAVAKP-----NTTTIOFPANQ 45
 DB 340 PNTAPPNFSHTGCTFPVPPFLSSIAGVPTTAATAPVATSPNDISTSVIGSEVTV 399
 QY 46 LPPGYLIKNSGFLMLVS---PQOTVTRAEFTSNITSRAVPANQVKTCT-----Y 96
 DB 400 TEEGIAGVATSTG--VVTSGGLPIPVSESPVLSVSSITIPA---VVSISTTSPSLQV 454
 QY 97 PNSSSOLIKK-----VAVTPVKKIAOIGTVVTVTPKSSVQSAVAPTSTVYTPCKPLN 151
 DB 455 PITSSELVSVSTALYSVYTSATSASAGSSTATPGRKPRPAAVVSQQAAGS---TTVATLTL 511
 QY 152 TVTTLKPSLIGASTPDSNEPLKAENSAVOINLSPMLENVKCKNFI-----AM 202
 DB 512 SVST-----TSPSTASOLSISLSSSTPS--FTLAETVVVASHSLDKTSHSSTTG 561
 QY 203 LKIL-ACSGSPENQNVKKLYEQILDARIEAEFTKLYELKSSPOHLPVPLKKS 261
 DB 562 AFSLSAPSSSSSPGAGVS-----SYISQPGCLHPLVP---SV 596
 QY 262 VALROLLPNSOSFTIOOCVOQOTSSDMV-----IATCTTTVTSPVTTVSSSOSEKSI 315
 DB 597 IASTPLP-----QAAGTSTPLLPQVPSITPLVQVPAVNAVQOITLHSPQALPL 649
 QY 316 -----VSGAPAPRVSVQTLNPLAGPVGAKAGVTLHSVGPAAATGGTTAGTGLQ 366
 DB 650 NQPHTCPEVSDTOPKAGCIDDKTLE-----EKLRSIFSEHSSGQAQHASVSL 700
 QY 367 TSKPLVTSVANVTYVLSQPEKPVVSGTAVTL--SLPAVTPEFTSGAAICLPSVAPVVS 423

```

Db 701 TSLVIESVTGPIPTTAVAPSKLLSTSTCLPPNLPGLT-----VALP-VTPVVT 751
QY 424 FCMWDHICKPVIGTPVQIKLA--QPG-----PVLISQAPIGTSSSKOLFSLF-- 468
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 752 -----PGQVSTPVSTTSGVKGPTAPASKPLTKAPVLPVGTLPAGTLPSEQLPPFPG 804
QY 469 ---HVQOPSGGNEKQVTTTISHSSTLIQCGQKTMVNTIIPTSQPPASILKQITLPG 525
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 805 PSLTOSQOPLLEDLDAQLRRRLTSPEMITV-----TSAVGPVSMAPTAITEAGTQPQ 855
QY 526 NKILSLQASPT-----OKNRKENVTSCEPDEDDIN---DVTSMAGVNLNEENAC 572
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 856 KGVSVQKEGPVLATSSGAGVFKMGRFOVSVADGAQKEGKNKSEDAKSVHFESSSTSESV 915
QY 573 ILATNSELVGTLIOCKDEPFLIGALQKRILIDIGKKHDITELNSDAVNLISOATOERLR 632
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 916 LSSSSPE--STLV---KPEP-----NGTITIGISSDVPESAHKTTASEAK 955
QY 633 GLEKRLAIQHRMTT-----YKASENYILCSDTRSQ-----LKLEKLDOL----- 674
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 956 SDTGQPTKVGRFQVTTANKVGRFSVSKTEDKITDTKKEGPVASEPFMD-LEOAVLPAYI 1014
QY 675 -EKQRKDLEREML-----LKAARS 695
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1015 PKKEPELSEPSHLNGPSSDEPAFLSRVDGSGSPHQSLSLPSONLSQSLSNS 1074
QY 696 -----NKEDEPOLRLKOKAKELQLELAQIOHRDAN----- 726
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1075 FNSGYMSDNESEDIEDLELRLRDKHLKEIODLOSROKHIESLYTKLGKVPYAVI 1134
QY 727 LTATAIGPRKKRPLE-----SGIEGLKDNLIASGTSSLATKQOLH 767
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1135 IPPAAPLGRRRRPTKSGSKSRSSSLGKNKSPQLSGNLSGQSAASVYLH 1183

```

Search completed: February 16, 2003, 21:58:41
 Job time : 81.8407 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:40:36 ; Search time 27.5208 Seconds
(without alignments)
1738.213 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_359

Perfect score: 1773
Sequence: 1 GRLVTRKAPVSAAPKVSQGP.....KAGVTLHSVGPATAGGTP 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1773	100.0	801	19 AAW31494	Human hTAFII105 pr
2	1773	100.0	801	22 ABC09468	Novel human diagno
3	1773	100.0	852	21 AAY57279	Transcription facto
4	491	27.7	737	15 AAR56494	TATA-binding prote
5	491	27.7	737	17 AAW6084	Human TATA-binding
6	491	27.7	737	18 AAW25019	Human TATA-binding
7	491	27.7	1023	23 AAU82954	Human homologue of
8	216	12.2	921	15 AAR56487	TATA-binding prote
9	216	12.2	921	17 AAW6077	Drosophila TATA-bl
10	216	12.2	921	18 AAW25028	TATA-binding prote

11	216	12.2	921	22 ABB61528	Drosophila melanog
12	216	12.2	921	22 ABB66055	Drosophila melanog
13	182	10.3	1795	22 ABB69806	Drosophila melanog
14	176	9.9	2035	15 AAR57141	Host cell factor p
15	175.5	9.9	708	22 AAW79978	Human protein SEQ
16	175.5	9.9	709	22 AAW78994	Human protein SEQ
17	175.5	9.9	709	22 AAB93063	Human protein sequ
18	175.5	9.9	709	22 AAB94334	Human protein sequ
19	175.5	9.9	881	22 AAM93811	Human polypeptide,
20	167.5	9.4	5179	22 AAM24516	C899P predicted am
21	164.5	9.3	752	15 AAR51701	AML1-MTG8 fusion.
22	159.5	9.0	557	22 AAB94078	Human protein sequ
23	159.5	9.0	2781	21 AAY57453	Human transcriptio
24	159.5	9.0	2907	21 AAY57452	Human transcriptio
25	156.5	8.8	1328	22 AAM78519	Human protein SEQ
26	156.5	8.8	1331	22 AAM79503	Human protein SEQ
27	156	8.8	1057	22 ABB65440	Drosophila melanog
28	156	8.8	1322	21 AAB42650	Human ORFX ORF2414
29	155.5	8.8	842	22 ABB66631	Drosophila melanog
30	155.5	8.8	864	22 ABB71319	Drosophila melanog
31	153	8.6	612	23 ABB62110	Human prostate spe
32	153	8.6	1296	23 ABB66702	Human novel polype
33	153	8.6	1296	23 ABB66756	Human novel polype
34	152.5	8.6	386	22 ABB29832	Peptide #2483 enco
35	152.5	8.6	386	22 ABB35008	Peptide #2514 enco
36	152.5	8.6	386	22 ABB20422	Protein #2421 enco
37	152.5	8.6	386	22 AAM55821	Human brain expres
38	152.5	8.6	386	22 AAM68195	Human bone marrow
39	152.5	8.6	386	22 AAM16015	Peptide #2449 enco
40	152.5	8.6	386	22 AAM28516	Peptide #2553 enco
41	152.5	8.6	386	22 AAM03749	Peptide #2431 enco
42	152.5	8.6	2870	21 AAY95559	Caenorhabditis ele
43	152.5	8.6	3178	21 AAY95556	Caenorhabditis ele
44	151	8.5	501	19 AAM82571	Human BBP1 DNA
45	150	8.5	571	22 AAB27242	Human EXMAD-20 SEQ

ALIGNMENTS

RESULT 1
AAW31494
ID AAW31494 standard; Protein; 801 AA.
AAW31494:
28-APR-1998 (first entry)
Human hTAFII105 protein.
TATA-binding protein associated binding factor 105; human; activator;
hTAFII105; transcription factor; TFIID; transcriptional activation;
antibodies; diagnosis; therapy; pharmaceutical industry.
Homo sapiens.
US5710025-A.
20-JAN-1998.
02-OCT-1996; 96US-0725012.
02-OCT-1996; 96US-0725012.
02-OCT-1996; 96US-0725012.
(REGC) UNIV CALIFORNIA.
Dikstein R, Tjian R;
WPI: 1998-109818/10.
N-PSDB; AAV02872.
DNA encoding human tata-binding protein associated factor - for
producing recombinant protein

XX Claim 1; Col 17-22; 12pp; English.
XX
XX This cDNA sequence represents a human tata-binding protein associated
CC factor, htaFII105, isolated from Daudi cell nuclear extracts. Tightly
CC associated subunits (TAF's) are components of the transcription factor
CC TFIID and are thought to mediate transcriptional activation. This encoded
CC protein may be produced recombinantly from transformed host cells or
CC purified from human cells. htaFII105 specific binding agents such as
CC specific antibodies could be used for diagnosis (e.g. genetic
CC hybridisation screens for htaFII105 transcripts), therapy (e.g. gene
CC therapy to modulate htaFII105 gene expression) and in the
CC biopharmaceutical industry (e.g. as immunogens, reagents for isolating B
CC cell specific activators or other transcriptional regulators).
XX
XX Sequence 801 AA;
SQ
Query Match 100.0%; Score 1773; DB 19; Length 801;
Best Local Similarity 100.0%; Pred. No. 7,8e-142;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTLVTKVAPVSAAPKVSQSGRLPAPQIVAVKABNTTITQEPANLQLPPTGVLIKNSGRL 60
DB 1 GTLVTKVAPVSAAPKVSQSGRLPAPQIVAVKABNTTITQEPANLQLPPTGVLIKNSGRL 60
QY 61 MLVSPQQTIVRAETTSNITSRPAVPANPQTIVKICTVPNSSQLIKVAVTPVKKLAQIGT 120
DB 61 MLVSPQQTIVRAETTSNITSRPAVPANPQTIVKICTVPNSSQLIKVAVTPVKKLAQIGT 120
QY 121 TVVTVTPKSSVQSVAVPTSVVTPGKPLNTVTLKPSLSIGASTPSNPNKAENSAA 180
DB 121 TVVTVTPKSSVQSVAVPTSVVTPGKPLNTVTLKPSLSIGASTPSNPNKAENSAA 180
QY 181 VOINLSPTMLENVKCKNPLAMLIKACSGSQSPKMGQNVKLVLEQLDAKIEAEFTTRK 240
DB 181 VOINLSPTMLENVKCKNPLAMLIKACSGSQSPKMGQNVKLVLEQLDAKIEAEFTTRK 240
QY 241 LVELEKSSPOPHLVPLPKKSVVALROLPLNSQSFIOQCVOQTSDDVIACTTTVTTSPV 300
DB 241 LVELEKSSPOPHLVPLPKKSVVALROLPLNSQSFIOQCVOQTSDDVIACTTTVTTSPV 300
QY 301 VTTTVSSOSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTVLHVSQPTAATGTT 359
DB 301 VTTTVSSOSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTVLHVSQPTAATGTT 359
RESULT 2
ABG09468
ID ABG09468 standard; Protein; 801 AA.
XX
XX ABG09468;
AC
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #9459.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Drmanac RT, Liu C, Tang YT;
PI

XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS73655.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID No 39827; 103pp; English.
PS
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 801 AA;
Query Match 100.0%; Score 1773; DB 22; Length 801;
Best Local Similarity 100.0%; Pred. No. 7,8e-142;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTLVTKVAPVSAAPKVSQSGRLPAPQIVAVKABNTTITQEPANLQLPPTGVLIKNSGRL 60
DB 1 GTLVTKVAPVSAAPKVSQSGRLPAPQIVAVKABNTTITQEPANLQLPPTGVLIKNSGRL 60
QY 61 MLVSPQQTIVRAETTSNITSRPAVPANPQTIVKICTVPNSSQLIKVAVTPVKKLAQIGT 120
DB 61 MLVSPQQTIVRAETTSNITSRPAVPANPQTIVKICTVPNSSQLIKVAVTPVKKLAQIGT 120
QY 121 TVVTVTPKSSVQSVAVPTSVVTPGKPLNTVTLKPSLSIGASTPSNPNKAENSAA 180
DB 121 TVVTVTPKSSVQSVAVPTSVVTPGKPLNTVTLKPSLSIGASTPSNPNKAENSAA 180
QY 181 VOINLSPTMLENVKCKNPLAMLIKACSGSQSPKMGQNVKLVLEQLDAKIEAEFTTRK 240
DB 181 VOINLSPTMLENVKCKNPLAMLIKACSGSQSPKMGQNVKLVLEQLDAKIEAEFTTRK 240
QY 241 LVELEKSSPOPHLVPLPKKSVVALROLPLNSQSFIOQCVOQTSDDVIACTTTVTTSPV 300
DB 241 LVELEKSSPOPHLVPLPKKSVVALROLPLNSQSFIOQCVOQTSDDVIACTTTVTTSPV 300
QY 301 VTTTVSSOSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTVLHVSQPTAATGTT 359
DB 301 VTTTVSSOSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTVLHVSQPTAATGTT 359
RESULT 3
AAY57279
ID AAY57279 standard; Protein; 852 AA.
XX
XX AAY57279;
AC
XX
XX 06-JUN-2000 (first entry)
DT
XX
XX Transcription factor subunit TAFII105 polypeptide.
DE

XX TATA box-binding protein associated factor II 105; TAFII105; cancer;
 KW transcription factor; apoptosis; cytosolic; immunosuppressive;
 KM antiinflammatory; virucide; antibacterial.
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 XX FH Misc-difference 802 /label= G1x
 FT FT /note= "encoded by TGA"
 FT FT /note= "encoded by TGA"
 FT FT /label= G1x
 FT FT /note= "encoded by TAA"
 XX MO200012699-A1.
 XX PD 09-MAR-2000.
 XX PF 25-AUG-1999; 99WO-IL00464.
 XX PR 27-AUG-1998; 98IL-0125971.
 XX PA (YEDA) YEDA RES & DEV CO LTD.
 XX PI Dikstein R, Yamit-hezi A;
 XX DI
 XX DR WPI: 2000-256640/22.
 XX N-PSDB; AA290465.
 XX PT Polypeptide encoding TATA box binding protein associated factor II 105
 PT useful for treating e.g. cancers and inducing apoptosis has a dominant
 PT negative effect on the normal biological activity of the binding
 PT protein.
 XX
 XX PS Claim 7; Fig 2; 48pp; English.
 XX
 CC This represents a polypeptide comprising a (modified) fragment (I) of
 CC a TATA box-binding protein associated factor II 105 (TAFII105). A
 CC pharmaceutical composition comprising (I) or the polynucleotide or an
 CC inhibitor or antagonist of (I) is useful for treating cancers and
 CC inducing apoptosis in pathological cells. The composition is also useful
 CC for treating autoimmune diseases, inflammatory processes and viral or
 CC bacterial infections.
 CC
 XX
 XX Sequence 852 AA:
 SQ
 Query Match 100.0%; Score 1773; DB 21; Length 852;
 Best Local Similarity 100.0%; Pred. No. 8.5e-142;
 Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 301 VTTVSSSQSEKSIIVSGATAPRTVSVOPLNPLAGPVGAKAGVTLHSGPFAATGTT 359
 RESULT 4
 ID AAR56494 standard; Protein; 737 AA.
 XX
 AC AAR56494;
 XX
 DT 23-MAR-1995 (first entry)
 XX
 XX TATA-binding protein-associated factor hTAFII130.
 DE
 XX TATA-binding protein associated factor; hTAFII130; screening;
 KW diagnostic; therapeutic; gene transcription regulation.
 KM
 XX Homo sapiens.
 OS
 XX MO9417087-A.
 XX PN 04-AUG-1994.
 XX PD 28-JAN-1994; 94WO-US01114.
 XX PF 28-JAN-1993; 93US-0013412;
 XX PR 30-JUN-1993; 93US-0087119.
 XX (REGC) UNIV CALIFORNIA.
 XX
 XX Comal L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
 XX Tjian R, Wang E, Weinzierl ROJ;
 XX
 XX WPI: 1994-264019/32.
 XX N-PSDB; AA070731.
 XX
 XX TATA-binding protein associated protein factors - and
 XX corresponding nucleotide sequence and deriv. antibodies, useful
 XX in screening, diagnostics and therapeutics
 XX
 PS Disclosure; Page 142; 180pp; English.
 XX
 CC The TATA-binding protein associated factor hTAFII130 (including
 CC specific antibodies and fusion products) are used in drug screening,
 CC diagnostics and therapeutics. They are used in the development of
 CC specific biochemical assays for screening compounds that agonise or
 CC antagonise selected transcription factors involved in regulating
 CC gene expression associated with human pathology.
 CC
 XX
 XX Sequence 737 AA;
 SQ
 Query Match 27.7%; Score 491; DB 15; Length 737;
 Best Local Similarity 38.4%; Pred. No. 6.9e-33;
 Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;
 QY 13 PPKVSSG-----PRLPAPQIVAVKAPNTTIOFPANLQLPQGVTLIKNSGPI 61
 DB 68 PLATTSGRATLTPTVLARLPQ-----PQNTFNIO---NFQPLPMTVLVRENGLL 118
 QY 62 LVSPQQTVTR-----AETSNITSRPAPNPQTVKICTVPSNSQLIKKVAATPVKKL 115
 DB 119 MI-PQALAOQAOAHQPPQTMAPRPTSPAPPVQISTVQAPGTPIAR-QVT----- 172
 QY 116 AQIGTIVTTPKPSVQSVAVPTSVTVTPGKPLNTYT--TLKPSLSGAS-----TPS 168
 DB 173 ---TTIIKQV---SQAQTVQPSATLQSPGVQPOLVLGGAQOTASLGTAVQGTGPQ 225
 QY 169 NE-PNLKAENSAVAOINLSPMLLENVKKCKNPLMLIKLACSGSQSPMGONVKLYEOL 227
 DB 226 RTVGGATITSSAATE-----TMENVKCKKNFLSTLIKLASGQSTETANAKVELYONL 279
 QY 228 LDKAIEAEETFRKLYVELKSSPPQHLVPELKKSVVALRQLLPNSQSFIOQCVOQ---TS 283
 DB 280 LDGKIEAEDEFTSRILYRELNSSPPQYLVPELKRSLPALRQLPPDAAAFIOCSQOQPPPTS 339

```

OY 284 SDWVATCTTWTSPVTTTSSOSEKSIIVSGATAPRTVS 327
DB 340 Q-----ATTALTAVLSSVQRTAGKTAATVTSALQPVLSL 376

RESULT 5
AAW06084
ID AAW06084 standard; Protein: 737 AA.
AC AAW06084;
XX
XX 27-JAN-1997 (first entry)
DE Human TATA-binding protein associated factor hTAFII130 protein.
XX
XX Human; TATA-binding protein; TBP associated factor; TFIID; holoenzyme;
XX RNA polymerase II; transcription; messenger RNA; nuclear fraction;
XX lambda-gt11; expression library.
OS Homo sapiens.
XX
XX US5534410-A.
XX
XX 09-JUL-1996.
XX
XX 28-JAN-1993; 93US-0013412.
XX
XX 28-JAN-1994; 94US-0188582.
XX 28-JAN-1993; 93US-0013412.
XX 30-JUN-1993; 93US-0087119.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Comai L, Dynlacht BD, Hoey T, Rupert S, Tanese N;
XX Tjian R, Wang E, Weinzierl ROJ;
XX
XX WPI: 1996-333245/33.
XX N-PSDB; AAT42217.
XX
XX Screen for cpds. that bind human TATA-binding protein associated
XX factor - by testing ability to bind to polypeptide fragments of the
XX factor, useful as (ant)agonists of transcription factors involved in
XX disease.
XX
XX Examples: Column 105-112; 86pp; English.
XX
XX This is the amino acid sequence of the human TATA-binding protein (TBP)
XX associated factor (TAF) designated TAFII130. The protein is a component
XX of the TFIID fraction required for reconstituting RNA polymerase II in
XX vitro transcription activity. The encoded protein has an estimated mol.
XX wt. of 130 kD by SDS-PAGE.
XX The invention relates to purified proteins involved in transcription
XX by RNA polymerase II, the RNA polymerase which transcribes messenger
XX RNA. RNA polymerase II transcription proceeds in vitro upon addition
XX of several nuclear fractions designated TFIIA, B, D, E, F, H, I and J
XX to RNA polymerase II holoenzyme. Fraction TFIID has been shown to
XX contain a TBP and other TAFs. Purification of TFIID and separation of
XX its components reveals 7 proteins ranging in size from 30-250 kD.
XX Serum raised against the TFIID fraction allowed cloning of the corresp.
XX genes from lambda-gt11 expression libraries.
XX
XX Sequence 737 AA:
XX
XX Query Match 27.7%; Score 491; DB 17; Length 737;
XX Best Local Similarity 38.4%; Pred. No. 6.9e-33;
XX Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;
OY 13 PKVSSG-----PRAPQVANKAKRTTITOPANIQLPFGVLLKSNGLM 61
DB 68 PTATTSIGIRATLTPVLARLP-----PONPTNIQ---NFOLEPGVAVLSENGQL 118
OY 62 LVSPQQTIVR-----AERTSNITSRAVAPANDQIVKICTVNSSQLKKVAVTPVKKL 115

```

```

DB 119 MT-PQAKLAQMAQAHQPTTMAPRPAPTSAAPVOISTVQAPGPILAR-OVTP----- 172
OY 116 AQIGTWTVTVPKPSVQSAVPTSVTVPGPLNTVT--TKPSSIGASS-----TPS 168
DB 173 -----TTIIKQV-----SQAGTIVQPSATLQGRSPGVQPOLVLOGAAGTASLGNATVQGTCP 225
OY 169 NE-PNLKAENSAVQINLSPTMLBNKKCKNFILAMLIKACSSQSEPMQNVKKTLEOT 227
DB 226 RTVPGATTTSSAATE-----TMENVKCKNFILSTLIKLASGQSTETAAVKEIVQNL 279
OY 228 LDAKIEAEFTRLVLEKSSPPHLPFLKRSVALROLPLPSQSFIOQCVQ-----TS 283
DB 280 LDKRIEAEFTSRLYRELNSSPPVLPFLKRSPLALROLTPDSALFIQDSQDQPPPPPS 339
OY 284 SDWVATCTTWTSPVTTTSSOSEKSIIVSGATAPRTVS 327
DB 340 Q-----ATTALTAVLSSVQRTAGKTAATVTSALQPVLSL 376

RESULT 6
AAW25019
ID AAW25019 standard; Protein: 737 AA.
XX
XX AAW25019;
XX
XX 08-OCT-1997 (first entry)
XX
XX TATA-binding protein associated factor, hTAFII130.
XX
XX TATA-binding protein associated factor; TAF; nuclear protein;
XX RNA polymerase transcription; TATA-binding protein; TBP;
XX Initiation.
XX
XX Homo sapiens.
XX
XX US5637686-A.
XX
XX 10-JUN-1997.
XX
XX 28-JAN-1993; 93US-0013412.
XX
XX 28-JAN-1994; 94US-0188582.
XX 28-JAN-1993; 93US-0013412.
XX 30-JUN-1993; 93US-0087119.
XX 09-MAY-1996; 96US-0646715.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Comai L, Dynlacht BD, Hoey T, Rupert S, Tanese N;
XX Tjian R, Wang E, Weinzierl ROJ;
XX
XX WPI: 1997-319113/29.
XX N-PSDB; AAT79595.
XX
XX Nucleic acids encoding human TATA-binding protein associated factor
XX (TAF) peptide(s) - for production of recombinant peptide(s), used
XX for modulating transcription of TAFs
XX
XX Claim 1; Column 111-116; 86pp; English.
XX
XX AAW25018 represents TATA-binding protein associated factor (TAF)
XX polypeptide, hTAFII100 (mol. weight 100kD). TAF peptides derived
XX from hTAFII30 alpha, hTAFII30 beta, hTAFII40, hTAFII70, hTAFII100,
XX hTAFII130, hTAFII250, hTAFII48 and hTAFII10 and nucleic acids encoding
XX them, are used to modulate transcription, including transcription
XX initiation. TAFs are nuclear proteins involved in RNA polymerase I,
XX II and III transcription. The peptides act by binding to a different
XX TAF, an activator, or TBP (TATA-binding protein) or competitively
XX inhibiting association of a TAF domain with another compound, typically
XX a protein like TBP or another TAF, an activator, or DNA.
XX
XX Sequence 737 AA:

```

Query Match	27.7%	Score 491;	DB 18;	Length 737;
Best Local Similarity	38.4%	Pred. No. 6.9e-33;		
Matches 132;	Conservative 52;	Mismatches 96;	Indels 64;	Gaps 14

```

QY 13 PPKXSSG-----PRLPAQIYAVKAPNTTITQFPAHQIPGTVLIXNSPLM 61
Db 68 PLATISGRATITFTVLAPRLPQP-----PQNTNIQ-----NFDLPBGWLVXSNQQL 118
QY 62 LVSPQOYVTR-----AETTSNITSRPAVPANPOTVAKICTVYPNSSODLIKAVATPVKYL 115
Db 119 MI-PQOALAQWQAQAHQPOPTMAPRATIPRSAPPOVISTYQACGTIILAR-QVTP----- 117Z
QY 116 AQISTVVTYVTKPSSVQSAVPTSVTVTPGKPLNTVT--TLKPSLXASS-----TPS 168
Db 173 -----TTIRKQV-----SQAQTVQVPSATLQSRSPQVQPLVLGGAAQTASLGTATAVQOTQPQ 225
QY 169 NE-PNLKAENSAVOINISPMLENVKKCKNFELMLIKLACSGSQSPEDMGONAKKLYEOL 222
Db 226 RTVGATTTSSAATE-----TKENYKKCKNFELSTLTIKLASSGQSFETATANYKELVQNL 279
QY 228 LDARIEAEFMRKLYVELKSPQCHLYLPELKKSVVALROLLPNQSFITQOCVQO-----TS 283
Db 280 LDGKIEADEFTSRILYRELNNSPQYLVLPFLKRSRLPALRQLTPDSEAAITQOSQOQPPPTTS 339
QY 284 SDMWIATCTTTSPTVYTVSSQSEKSIYVGATAPRAYS 327
Db 340 Q-----ATTALTAVALVSSVQRTACKTATATVTSALQOPVLSL 376

```

RESULT 7
AAU82954
ID AU82954 standard; Protein; 1023 AA

AC AAU82954 ;

DT 23-APR-2002 (first entry)

Human homologue of MPT1 protein target for antifungal compound.

KM AftIIfungus; fungal gene transcription; RPC34; POD3; TFN2; NAB2;
KM MPT1; MTR2; BOS1; POL30; KSA2; SQT1; MTW1; SPG98; BFE2; RNA1;
KM GCD7; SKR6; NIP1; LCP5; NCE103; ECOL; ORC2; CNS1; YPD1; TIM10; SBA4;
KM yeast; fungus.

OS Homo sapiens.

PN W0200202055-A2.
XY

PD 10-JAN-2002.
XX

Pf 28-JUN-2001; 2001WO-US20592.
XX

PR 29-JUN-2000; 2000US-213164P.
PR 10-AUG-2000; 2000US-224457P.

PA (ANAD-) ANADYS PHARM INC.

PI Moore J, Buurman ET, De

PI Davidov E, Thompson CM;

DR WPI: 2002-147962/19.

XX
CM

PT protein, modulation of essential protein activity, binding to essential protein by contacting protein with test compound, and determining

pt effects -
XX

PS Claim 1; figure 9; 524pp; English.
XX

CC compound/or interaction with essential proteins (EP) or for modulation
CC of EP activity e.g fungal gene transcription. The proteins tested in the
CC invention include RPEC34, P033, TFA2, NAB2, MPT1, MPR2, BOS1, POL30, R5A2,
CC SOT1, MTW1, TFB1, SPX98, RPF2, RNL1, GCD7, SK16, N1P1, ICG55, NCE103,
CC ECOL1, ORC2, CNS1, YP11, TIM10 and SR44 from *S. cerevisiae*, *C. albicans*
CC and human homologues. The method involves contacting a culture with one
CC or more test compounds and determining the effects on the growth or
CC viability of the culture of cells which preferably comprises fungal cells
CC or yeast cells. Preferably the identified compounds interact with, or
CC modulate (preferably inhibit) activity of *C. albicans* EP. The inhibitor
CC compounds identified by the method are useful for preventing or
CC inhibiting fungal, particularly *C. albicans* growth in culture or in a
CC mammal. The antifungal agents interact with essential fungal elements
CC that can be used to treat fungal infection by preventing the growth and
CC preferentially killing the fungi, but does not inhibit the biological
CC activity of mammalian homologues. This amino acid sequence represents a
CC target protein used to test the antifungal compounds, described in the
CC method of the invention.

sq	sequence	1023 AA
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100
101	101	101
102	102	102
103	103	103
104	104	104
105	105	105
106	106	106
107	107	107
108	108	108
109	109	109
110	110	110
111	111	111
112	112	112
113	113	113
114	114	114
115	115	115
116	116	116
117	117	117
118	118	118
119	119	119
120	120	120
121	121	121
122	122	122
123	123	123
124	124	124
125	125	125
126	126	126
127	127	127
128	128	128
129	129	

Query Match	27.7%;	Score 491;	DB 23;	Length 1023;
Best Local Similarity	38.4%;	Pred. No. 1,1e-32;		
Matches 132;	Conservative 52;	Mismatches 96;	Indels 64;	Gaps 14;

```

0Y 13 PKPVSSG-----PRJAPDIVAKPNTTIOFANLLOPOTVIEKNSBIM 61
   | : : : | | | | | : | | | | | : | | | | : | | | |
Db 414 PRATTSGRILTLPTVLAPRLPQ-----PQNTNIO---NFOUPRGMVLVSENGOL 466
   | : : : | | | | | : | | | | | : | | | | : | | | |
0Y 62 LVSPQOQYTR-----AETTSNITSRPANVPANPTQVICKIYPNSSSOLIKKVAATPYKTL 115
   : : : : | : : : | : : : | : : : | : : : | : : : |
Db 465 MI-PQOALAQOQAQAHQPOQTMAPRATPTISAPRYOSTYQABGTPIAR-QYTP----- 511
   : : : : | : : : | : : : | : : : | : : : | : : : |
0Y 116 AOIGTTVTYVPRBSSVQSAVPVTSVTYTPGKPLNTYT-TLKPSSIGASS-----TPS 166
   : : : : | : : : | : : : | : : : | : : : | : : : |
Db 519 ----TTIKQV----SQAOITVQPSATLQTRSGVQPOLVLGGAQATASIGRTAVQGTGPQ 573
   : : : : | : : : | : : : | : : : | : : : | : : : |
0Y 169 NE-PLUKKENSANAOINISPTMLEVYKCKKFLMLIKLACSSGQSPBMONKRYLYBOL 222
   : : : : | : : : | : : : | : : : | : : : | : : : |
Db 572 RTVPATTTSSAYTE-----TMEVYKCKKFNLESTLLIKLASSGQSTETANVLELYONT 625
   : : : : | : : : | : : : | : : : | : : : | : : : |
0Y 228 LDKATEAEFTPRKLYVELKSSPOPLVETLKKSVALROLLPNRSOFIOOCVOO-----TS 283
   | : : : : | : : | | : : : | : : : | : : : | : : : |
Db 626 LDGKTEADEFTSRILRELNNSSQPLVPEFLKRSIPALMQPLPDSAAFIQOOSQOQPPPTTS 688
   : : : : | : : | | : : : | : : : | : : : | : : : |
0Y 284 SDMVATICTTYYTSPVVTYVSSQSOSEKSIIVSGATAPRTYVS 327
   | : : : | : : : | : : : | : : : | : : : | : : : |
Db 666 Q-----ATTALTAVLVSSSQVQKAGTAAITVTSALQPVLSL 722

```

RESULT 8	
AAR56487	
ID AAR56487	standard; Protein; 921 AA

AC AAR56487;
VV

DT 23-MAR-1995 (first entry)
YY

DE TATA-binding protein-associated factor dTAFII110.
XX

KW diagnostic; therapeutic; gene transcription regulation.

05 *Drosophila*.

PN W09417087-A

PD 04-AUG-1994

PF 28-JAN-1994; 94WO-US01114.

PR 28-JAN-1993; 93US-0013412.

XX (REGC) UNIV CALIFORNIA.
 XX PA
 XX PI Comal L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
 XX PI Tjian R, Wang E, Weinzierl ROJ;
 XX DR MPI: 1994-264019/32.
 DR N-PSDB: AAO70724.
 XX
 XX PT TATA-binding protein associated protein factors - and
 PT corresponding nucleotide sequence and deriv. antibodies, useful
 PT in screening, diagnostics and therapeutics
 XX
 XX PS Disclosure; Page 56-61; 180pp; English.
 XX
 CC The TATA-binding protein associated factor dTAFII110 (including
 CC specific antibodies and fusion products) are used in drug screening,
 CC diagnostics and therapeutics. They are used in the development of
 CC specific biochemical assays for screening compounds that agonise or
 CC antagonise selected transcription factors involved in regulating
 CC gene expression associated with human pathology.
 CC
 XX
 SQ Sequence 921 AA:
 Query Match 12.2%; Score 216; DB 15; Length 921;
 Best Local Similarity 24.9%; Pred. No. 2.3e-09;
 Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;
 QY 25 POIVAAKAPNTTITOPANLQLPPTGTVLIRKNSGRLML-----VSPQQTVAETTS 76
 DB 133 POSPTITLSTLNTGQTPA-----LVKTDNGFOLLRGTTGTPPTVQITTTNNSNS 184
 QY 77 NITSRAVPANPQTVKICIVPNSSQ-----LIKVAATPVKKLAQIGTTVTTVP 127
 DB 185 NTSTNNHPTTQ-IRLQTVPAASMTNTATSNITIVNSVASSGVANSSQPHLTLQLMNQ 243
 QY 128 KPSSVQSVAVPTSVTVTEKPLNTVTTLKPSLSGASTPSNEPNLKAENSAVQNLSP 187
 DB 244 AQLPQITQITIPAQSOQOQVNNVSSAGTATAVSSTA-----ATT 287
 QY 188 TMLENVK-KCKNFLAMLIKACSGSOPMGQVKKVLBQLDAKTEAEFTKLYVEIK 246
 DB 288 TOGNTKCKRFLANLIEL--STREPKVEKRVRLIOELVANVNEPEEFCRLRLIN 345
 QY 247 SSPQHLVPELKKSVVALROL-----LPNSQSFQ--- 276
 DB 346 ASPQCLIGFLKSLPLRLQALYTKELVTEGIKPPQHVGLAGLSQQLPKIOAQIRPIG 405
 QY 277 ---OCVOQTSSDMVATCTTTVTTSPTVTTVSSQSEKSIIVSGATAPRTVS---VOT 329
 DB 406 PSQTTTIGQTVAMI--TFNALGTPTPTIGHTTISKQPN---IRLPAPRLVNTGIRI 460
 QY 330 LNPAGPVGAKAGVVTLHSGVPTAA 354
 DB 461 QIP-LOYVGQANIVQIR--GFOHA 482
 RESULT 9
 AAM06077
 ID AAM06077 standard; Protein; 921 AA.
 XX
 XX AC AAM06077;
 XX
 XX 27-JAN-1997 (first entry)
 DE Drosophila TATA-binding protein associated factor dTAFII110 protein.
 XX
 KW Drosophila TATA-binding protein; TBP associated factor; TFIID;
 KW RNA polymerase II; transcription; messenger RNA; nuclear fraction;
 KW holoenzyme; lambda-gli1; expression library.
 XX
 OS Drosophila melanogaster.
 XX

PN US5534410-A.
 XX PD 09-JUL-1996.
 XX PF 28-JAN-1993; 93US-0013412.
 XX PR 28-JAN-1994; 94US-0188582.
 PR 28-JAN-1993; 93US-0013412.
 PR 30-JUN-1993; 93US-0087119.
 XX
 XX PA (REGC) UNIV CALIFORNIA.
 XX PI Comal L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
 XX PI Tjian R, Wang E, Weinzierl ROJ;
 XX DR MPI: 1996-333245/33.
 DR N-PSDB: AAT42210.
 XX
 XX PT Screen for cpds. that bind human TATA-binding protein associated
 PT factor - by testing ability to bind to polypeptide fragments of the
 PT factor, useful as (ant)agonists of transcription factors involved in
 PT disease.
 XX
 XX PS Examples; Column 27-36; 86pp; English.
 CC This is the amino acid sequence of the Drosophila TATA-binding protein
 CC (TBP) associated factor (TAF) designated TAFII110. The protein is a
 CC component of the TFIID fraction required for reconstituting RNA
 CC polymerase II in vitro transcription activity. The encoded protein
 CC has an estimated mol. wt. of 110 kD by SDS-PAGE, a calculated mol. wt.
 CC based on sequence of the gene of 99.4 kD and an estimated pI of 10.1.
 CC The invention relates to purified proteins involved in transcription
 CC by RNA polymerase II, the RNA polymerase which transcribes messenger
 CC RNA. RNA polymerase II transcription proceeds in vitro upon addition
 CC of several nuclear fractions designated TFIIA, B, D, E, F, H, I and J
 CC to RNA polymerase II holoenzyme. Fraction TFIID has been shown to
 CC contain a TBP and other TAFs. Purification of TFIID and separation of
 CC its components reveals 7 proteins ranging in size from 30-250 kD.
 CC Serum raised against the TFIID fraction allowed cloning of the corresp.
 CC genes from lambda-gli1 expression libraries.
 CC
 XX
 SQ Sequence 921 AA:
 Query Match 12.2%; Score 216; DB 17; Length 921;
 Best Local Similarity 24.9%; Pred. No. 2.3e-09;
 Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;
 QY 25 POIVAAKAPNTTITOPANLQLPPTGTVLIRKNSGRLML-----VSPQQTVAETTS 76
 DB 133 POSPTITLSTLNTGQTPA-----LVKTDNGFOLLRGTTGTPPTVQITTTNNSNS 184
 QY 77 NITSRAVPANPQTVKICIVPNSSQ-----LIKVAATPVKKLAQIGTTVTTVP 127
 DB 185 NTSTNNHPTTQ-IRLQTVPAASMTNTATSNITIVNSVASSGVANSSQPHLTLQLMNQ 243
 QY 128 KPSSVQSVAVPTSVTVTEKPLNTVTTLKPSLSGASTPSNEPNLKAENSAVQNLSP 187
 DB 244 AQLPQITQITIPAQSOQOQVNNVSSAGTATAVSSTA-----ATT 287
 QY 247 SSPQHLVPELKKSVVALROL-----LPNSQSFQ--- 276
 DB 346 ASPQCLIGFLKSLPLRLQALYTKELVTEGIKPPQHVGLAGLSQQLPKIOAQIRPIG 405
 QY 277 ---OCVOQTSSDMVATCTTTVTTSPTVTTVSSQSEKSIIVSGATAPRTVS---VOT 329
 DB 406 PSQTTTIGQTVAMI--TFNALGTPTPTIGHTTISKQPN---IRLPAPRLVNTGIRI 460
 QY 330 LNPAGPVGAKAGVVTLHSGVPTAA 354

Db 461 QIP-SLOVPGQANIVQIR--GPOHA 482

RESULT 10
AAW25028
ID AAW25028 standard; Protein; 921 AA.
XX
AC AAW25028;
XX
DT 08-OCT-1997 (first entry)
XX
DE TATA-binding protein associated factor, dTAFl110.
XX
KW TATA-binding protein associated factor; TAF; nuclear protein;
KM RNA polymerase transcription; TATA-binding protein; TBP;
XX Initiation.
XX
OS Drosophila sp.
XX
PN US5637686-A.
XX
PD 10-JUN-1997.
XX
PF 28-JAN-1993; 93US-0013412.
XX
PR 28-JAN-1994; 94US-0188582.
XX 28-JAN-1993; 93US-0013412.
PR 30-JUN-1993; 93US-0087119.
PR 09-MAY-1996; 96US-0646715.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Comal L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
PI Tjian R, Wang E, Weinzierl ROJ;
XX
DR WPI; 1997-319113/29.
XX N-PSDB; AAT79604.
XX
PT Nucleic acids encoding human TATA-binding protein associated factor
PT (TAF) peptide(s) - for production of recombinant peptide(s), used
PT for modulating transcription of TAFs
XX
PS Example 1; Column 35-40; 86pp; English.
XX
CC AAW25028 represents TATA-binding protein associated factor (TAF)
CC polypeptide, dTAFl110 (mol. weight 110KD). TAF peptides derived
CC from dTAFl110 alpha, dTAFl110 beta, dTAFl140, dTAFl160, dTAFl180,
CC dTAFl1110, dTAFl150, and dTAFl1250, their human equivalents and
CC nucleic acids encoding them, are used to modulate transcription,
CC including transcription initiation. TAFs are nuclear proteins involved
CC in RNA polymerase I, II and III transcription. The peptides act by
CC binding to a different TAF, an activator, or TBP (TATA-binding protein)
CC or competitively inhibiting association of a TAF domain with another
CC compound, typically a protein like TBP or another TAF, an activator,
CC or DNA.
XX
SQ Sequence 921 AA;

Query Match 12.2%; Score 216; DB 18; Length 921;
Best Local Similarity 24.9%; Pred. No. 2.3e-09;
Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;

QY 25 PQIYAVKAPNTTITPQANLQLPQTVLIRKNSGPIML-----VSPQQTVAERTS 76
DB 133 PQSSISLTSLNTIGQTPA-----LLVKDKNFQOLLRVTTGGPPVYTIINTSINS 184
QY 77 NITSRPVAPNPOTVTKICTVPPNSSQ-----LIKKAATVPYKKLAQIGTTVVTVP 127
DB 185 NTSTSTNHPTTQO-IRLQTVPAASMTNTATSNIIYNSVASSGVANSOPHLTQLNAQ 243
QY 128 KPSSVGVAVPTSVYVTPGKPLNTVTTLRKSSLSGASTPSENENLRKRNANAVQINISF 187
DB 244 APQLPQITQIOTIPAGOSQOOOVNNSVAGGTAAVASTTA-----ATT 287

QY 188 TMLENVK-KCKNFLAMLIKACSGSPDMGONVAKLEOULDAKIEAEFEFRKLYEIK 246
DB 288 TQGGTKKCKRKFLLANLEL---STREPKVKEKNVTLQELVNAVNEDEFCDRERLLN 345
QY 247 SSPQPHLVPLKKSVALRQL-----LPSQSFEQ--- 276
DB 346 ASPQCLIGFLKKSILPLLRLQALYTKELVIEGKPPQVHLGAGSQOLPKIOAQIRPIG 405
QY 277 ----QCVQGTSSDWIACITTTVTSPVYTTVSSQSEKSIYSGATAPRVS---VQT 329
DB 406 PSQTTTIGQTVRMT--TPNALGTPRPRTGHTTISKQPN--IRLPAPRLVNTGIRT 460
QY 330 LNPAGPYGAKAGVYTLHSGPTAA 354
DB 461 QIP-SLOVPGQANIVQIR--GPOHA 482

RESULT 11
ABB61528
ID ABB61528 standard; Protein; 921 AA.
XX
AC ABB61528;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 11376.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX N-PSDB; ABL05631.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 11376; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences and pharmaceutical drugs, expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB5737-ABB72072).
CC
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 921 AA;

Query Match 12.2%; Score 216; DB 22; Length 921;
Best Local Similarity 24.9%; Pred. No. 2.3e-09;
Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;

QY 25 POIYAVKAPNTTIOFPANLQLPFGTVLIRKNSGRLML-----VSPOQVTRAEFTS 76
 DB 133 POSPSTLTLTNTGQTPA-----LVKTDNGFOLLRGVTTGTPPTVQTITNTSNNS 184
 QY 77 NITSRAVPANPQTVKICIVPNSSQ-----LIKVAVTPVKKLAQIGTVVTVTP 127
 DB 185 NTSTSTNHPTTQ-IRLQTVPAASMTNTATSNIIYNSVASSGYANSSQPHILQIQA 243
 QY 128 KPSVGSVAVPVTSVVTYTPKPLNTVTTLKPSLSGASSTPSNEPNLKAENSAVAQINLSP 187
 DB 244 APOLPQITQIOTIPAOOSQOQOVNNSVAGGTATVAVSSTTA-----ATT 287
 QY 188 TMLENVK-KCKNFLAMIKLACSGSOSPEMGONVKVLEOLDLAKTEAEFTKLYELK 246
 DB 288 TQGNTEKCKRKLNLIEL--STREPKVEKNVRLIOELVNAVNEPEEFCORLERLN 345
 QY 247 SSPQHLVPELKKSVVALROL-----LPNSQSFID-- 276
 DB 346 ASPQCLIGFLKSLPLRLQALYTKELVIGIKRPPQHVLAGLSQQLPKIQAQIRPIG 405
 QY 277 ----QCVOQTSSDMVATCTTCTTCTTSPVVTYVSSSOSSEKSTIVSGATAPRTVS---VOT 329
 DB 406 PSQTTTIGQTVRMI--TPNALGTPRTTIGTHTTISKQPPN--IRLPTARLVNTGIRP 460
 QY 330 LNPAGPVGAKAGVTLHSGVPTAA 354
 DB 461 QIP-SIQVPGQANIVQIR--GPOHA 482

RESULT 12

ABB66055 standard; Protein: 921 AA.

26-MAR-2002 (first entry)

Drosophila melanogaster polypeptide SEQ ID NO 24957.

Drosophila: developmental biology; cell signalling; insecticide; pharmaceutical.

Drosophila melanogaster.

WO200171042-A2.

23-MAR-2001; 2001WO-US09231.

23-MAR-2000; 2000US-191637P.
11-JUL-2000; 2000US-0614150.

(PEKE) PE CORP NY.

Venter JC, Adams M, Li PMD, Myers EW;

WPI; 2001-656860/75.
N-PSDB; ABL10158.

New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

Disclosure; SEQ ID NO 24957; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 SQ Sequence 921 AA:

Query Match 12.2%; Score 216; DB 22; Length 921;
 Best Local Similarity 24.9%; Pred. No. 2.3e-09;
 Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;

QY 25 POIYAVKAPNTTIOFPANLQLPFGTVLIRKNSGRLML-----VSPOQVTRAEFTS 76
 DB 133 POSPSTLTLTNTGQTPA-----LVKTDNGFOLLRGVTTGTPPTVQTITNTSNNS 184
 QY 77 NITSRAVPANPQTVKICIVPNSSQ-----LIKVAVTPVKKLAQIGTVVTVTP 127
 DB 185 NTSTSTNHPTTQ-IRLQTVPAASMTNTATSNIIYNSVASSGYANSSQPHILQIQA 243
 QY 128 KPSVGSVAVPVTSVVTYTPKPLNTVTTLKPSLSGASSTPSNEPNLKAENSAVAQINLSP 187
 DB 244 APOLPQITQIOTIPAOOSQOQOVNNSVAGGTATVAVSSTTA-----ATT 287
 QY 188 TMLENVK-KCKNFLAMIKLACSGSOSPEMGONVKVLEOLDLAKTEAEFTKLYELK 246
 DB 288 TQGNTEKCKRKLNLIEL--STREPKVEKNVRLIOELVNAVNEPEEFCORLERLN 345
 QY 247 SSPQHLVPELKKSVVALROL-----LPNSQSFID-- 276
 DB 346 ASPQCLIGFLKSLPLRLQALYTKELVIGIKRPPQHVLAGLSQQLPKIQAQIRPIG 405
 QY 277 ----QCVOQTSSDMVATCTTCTTCTTSPVVTYVSSSOSSEKSTIVSGATAPRTVS---VOT 329
 DB 406 PSQTTTIGQTVRMI--TPNALGTPRTTIGTHTTISKQPPN--IRLPTARLVNTGIRP 460
 QY 330 LNPAGPVGAKAGVTLHSGVPTAA 354
 DB 461 QIP-SIQVPGQANIVQIR--GPOHA 482

RESULT 13

ABB69806 standard; Protein: 1795 AA.

26-MAR-2002 (first entry)

Drosophila melanogaster polypeptide SEQ ID NO 36210.

Drosophila: developmental biology; cell signalling; insecticide; pharmaceutical.

Drosophila melanogaster.

WO200171042-A2.

23-MAR-2001; 2001WO-US09231.

23-MAR-2000; 2000US-191637P.
11-JUL-2000; 2000US-0614150.

(PEKE) PE CORP NY.

Venter JC, Adams M, Li PMD, Myers EW;

WPI; 2001-656860/75.
N-PSDB; ABL13909.

New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -
XX
PS Disclosure; SEQ ID NO 36210; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).
XX
SQ Sequence 1795 AA;

Query Match 10.3%; Score 182; DB 22; Length 1795;
Best Local Similarity 22.1%; Pred. No. 4.7e-06;
Matches 86; Conservative 52; Mismatches 165; Indels 86; Gaps 13;

QY 9 PVSAAPKVSQPRLPAPQIVAVKAPNTTITQIPANLQI-----PGTVLIKNSG 58
DB 700 PPSSTGKRTTPK-PSRTTPTTKVTITTTITPRLSSTETTSPTPTTTPQPTT 758
QY 59 PLMLVSPQOVYRAETTSNITSRAVPANPQTVKICIVPNSSSOLIKKVAVTPVKILQI 118
DB 759 TLLVTPKSTTTTTEKRTSSPK-PTTQTKTSTAPNTT-----KVALTTQETPT 812
QY 119 GTT-----VTVTPKPS-----VQSAVPTSVTVTPCKPLNT 152
DB 813 QSTSTPTTTRKTTNNPEPTSTKRTISTPKPSTTPKSTVASSTKRTISSPKPTE 872
QY 153 VTTLKPSISGASPTSPNPNKKAENSAVOINLSPTMLNKKCNFLAMLIKLAGSQ 212
DB 873 KSTENPTNSKTSALTSSTORA-----ISTSEPTKTQNTTTTPKPTTKTS 922
QY 213 SPEGQNVKLVLEQLDAK--TEAEFTRLKLVLEKSSPOHVLPELKKSVVALROLDPN 270
DB 923 TQEAATTSQKSVTVITTKKATESPLTTLSTEENTPPKP-----LRTTPT 970
QY 271 SOSITQOCVOQTSSDMVATCTTVTTSPTVTVSSQSEKSIIVGSA--TAPR--TVS 326
DB 971 TTS-----VYATRTITTTTISESSTETTSQKPKSTPTSTRTTPKVTVI 1017
QY 327 VQTLNPLAGPVGAKAGVTVLHSGPTAAT 355
DB 1018 VSTQNPPT--TTSKTSVTTTTPNPSPT 1044

RESULT 14
AAR57141
ID AAR57141 standard; Protein; 2035 AA.
XX
AC AAR57141;
XX
DT 19-MAR-1995 (first entry)
XX
DE Host cell factor protein.
XX
KW Herpes simplex virus; herpes virus; VP16; immediate early gene;
KW host cell factor; virus infection therapy; cellular protein;
KW transcription.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Active-site 21..31
FT Active-site /note= "peptide R60"
FT Active-site 168..186
FT Active-site /note= "peptide R37"
FT Active-site 333..340

FT Active-site /note= "peptide R52"
FT 426..449
FT /note= "peptide 362"
FT 511..526
FT /note= "peptide 329"
FT 578..594
FT /note= "peptide 223 first sequence"
FT 594..611
FT /note= "peptide R26 1st peptide"
FT 611..623
FT /note= "peptide 223 2nd sequence"
FT 723..731
FT /note= "peptide 318"
FT 802..813
FT /note= "peptide 299"
FT 813..820
FT /note= "peptide 268"
FT 836..847
FT /note= "peptide R26 2nd sequence"
FT 1010..1031
FT /note= "THE TNT repeat 1"
FT 1072..1093
FT /note= "THE TNT repeat 2"
FT 1101..1126
FT /note= "THE TNT repeat 3"
FT 1158..1183
FT /note= "THE TNT repeat 4"
FT 1286..1311
FT /note= "THE TNT repeat 5"
FT 1314..1339
FT /note= "THE TNT repeat 6"
FT 1349..1374
FT /note= "THE TNT repeat 7"
FT 1414..1439
FT /note= "THE TNT repeat 8"
FT 1774..1781
FT /note= "peptide 293 2nd sequence"
FT 1808..1819
FT /note= "peptide 115"
FT 1819..1840
FT /note= "peptide 261 1st sequence"
FT 1853..1863
FT /note= "peptide 240"
FT 1901..1919
FT /note= "peptide R32"
FT 1919..1930
FT /note= "peptide 261 2nd sequence"
XX
XX W09413315-A.
XX
XX 23-JUN-1994.
XX
XX 03-DEC-1993; 93WO-US11721.
XX
XX 04-DEC-1992; 92US-0989842.
XX 12-APR-1993; 93US-0046585.
XX
XX (GOLD-) COLD SPRING HARBOR LAB.
XX (TULARIX) TULARIX INC.
XX
XX Herr W, Lamarco K, Wilson A;
XX
XX WPI; 1994-234207/28.
XX N-PSDB; AAQ69229.
XX
XX New Host Cell Factor polypeptide(s) and nucleic acid - are used
XX to develop agents for diagnosis or treatment of disease
XX associated with expression of a HCF-modulated gene e.g. viral
XX infections
XX
XX Disclosure; Page 39; 71pp; English.
XX
XX HCF is required for the transcription of a number of
CC

CC viral genes, such as the immediate early herpes simplex virus-1
CC genes. Epitopes of the encoded protein can be used in
CC defining functional domains of HCF, identifying compounds that
CC associate with HCF or designing compounds capable of modifying HCF
CC transcription. Such agents can be used to treat viral infections.

SO Sequence 2035 AA:

Query Match 9.9%; Score 176; DB 15; Length 2035;
Best Local Similarity 24.2%; Pred. No. 1.8e-05;
Matches 108; Conservative 57; Mismatches 148; Indels 134; Gaps 22;

QY 10 VSAPRPVSSGRLPAQIYAVKAPNTTITOPFANLQPEPTVLKISGCLMLVSPQQT 69
DB 537 IGSSPQMSGMAALAAAAATQKIPSSA---PVLVSPACTTIVKT---MAVTPQT 588
QY 70 TRAEITSNITSRPAVPANPOT-----VKICTVPPSSSO---LIKKVAVTPYK 115
DB 589 LPA--TVKVAASPVWWSNPATRLKTAAGVGSVSSATPTSTRPITTVHKSCTVVAQ 646
QY 116 AQTGTTV-----TVPKP-----SSVQAVAPTS----- 140
DB 647 AGVTTVGVGVTKITLVKSPISVPGSALISMLGKMSVYQKPVQTSVAVTQASTNGPV 706
QY 141 -----VTVTPGKPLNTVTT-----LKPSGLASSST-----PSNE 170
DB 707 TQIIQKGPLPACTILKLVTSADGKPTTITQASGAGTKPYLLGSSVSPSTTKRGT 766
QY 171 PNKAENSAVAQINLSPMLNENKCKCNFLAML--IKLACSGSOSPENQAVKRLVEQLD 229
DB 767 TIKTIPMSLITQAGATGVTSSPGIKSPITITTKVMTSGTAP-----AKIITAV-- 818
QY 230 AKIEAEETFRKL-YVELKSSP-QP---HLVPLKRSVALROLPLNSQSFIOCVQOTS 283
DB 819 PKATGHHGQGVYQVVLKGAPOGPTILRTVP-----MGVRLVTPYVSAVKPAV---- 869
QY 284 SDWVIAFCTTPTVTSPVYTTVSSSOS-----EKSIIVSGATFARTVSVQTLNP 332
DB 870 -TTLVVKGTGTVTLGTVSTISLAGAGHSTASLATPITTLGITA--TLSSOVINP 926
QY 333 LAGPVGAKAGVTLHSGVPAATAGTT 359
DB 927 TA--ITVSAAGTTL-----TAAGGLTT 946

RESULT 15
AAM79978
ID AAM79978 standard; Protein; 708 AA.
XX
AC AAM79978;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 3624.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001MO-US04098.
XX
PR 03-FEB-2000; 2000US-0496814.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.

PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
DR WPI: 2001-476283/51.
DR N-PSDB; AAK53111.

CC Nucleic acids encoding polypeptides with cytokine-like activities,
CC useful in diagnosis and gene therapy -
PS Claim 20; Page 401; 6221pp; English.

CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.

SO Sequence 708 AA:

Query Match 9.9%; Score 175.5; DB 22; Length 708;
Best Local Similarity 25.4%; Pred. No. 4.4e-06;
Matches 106; Conservative 53; Mismatches 166; Indels 93; Gaps 20;

QY 3 LTVKVAVPSAPPK-----VSSGRLPAQIYAVKAPNTTIT---QEPAN----- 43
DB 239 LVTPRAVAVSSQPKLQRPVTSVSG-SLTATSV--LAPNATVAVATQVYSGNRPQTSLQPL 295
QY 44 -----LQLEPGTVLKSNSGRLMLVSPQQTVRAETTSNITSRPAVPAN 87
DB 296 PVILHNPVAVSSQPOLQSHRGLVYTNQPSGNEFISVQSPPTVSGLTKNPVSLPSLP-N 354
QY 88 PQTVKICTVNSSQLKKKAAVTPVKLAQIGTTV--VTYPRKSSVQSVAVPVSVTYTP 146
DB 355 P--TRNNVPSVPSPIQR--NPTASAAPLGTTLAAOVAAPTAHSIVQ-ATRTSLPTVGP 408
QY 147 GKPLNTVTLKP-----SSLGASSTPSNEPMLKENSAAVQINLSPMLEVKK 195
DB 409 SGLVSPSTNGPIQMKRIPISAFSTSSAAEQNSNTPTRIENQTKTIDASYSKKAASSTQ 468
QY 196 CKNFLAMLKILACSGSOS-----PMG--QNVKRLVEQLDLAKIEAEETFRKLYV 243
DB 469 CGK-----ATGSDSSGVIDLTMDEESGASQDKKLNHPVSGVSSSQVSRPLQ 519
QY 244 ELKSSP-QPHLVPLFKRSVALROLPLNSQSFIOCVQOQSSDMVATCTTYYTTSV--- 299
DB 520 IQPAPPLQPSGVPISGSGQTTI-HLPTAPPT---TVNVTHRPVTOVTRRLPVPRAVANH 574
QY 300 -VYTTVSSSSQSEKSI--IVSGATAPRTVSVQTLNPLAGVGAAGV---VTLHSGVP 351
DB 575 QVYVYTTLPAPPAQAPLNGCTVMAQAPAVAGVAPQNSVTVRVQTTTYYVNNGLTIGSGCP 632

Search completed: February 16, 2003, 21:55:06
Job time : 34.5208 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:59:29 ; Search time 9.22484 Seconds
(without alignments)
1145.042 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_359

Perfect score: 1773
Sequence: 1 GTLVTKVAPVAPKVSQGP.....KAGVTLHSVGTAAATGTT 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5a_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5b_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6a_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6b_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backlitest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1773	100.0	801	1 US-08-725-012-2	Sequence 2, Appl
2	491	27.7	737	1 US-08-188-582-16	Sequence 16, Appl
3	491	27.7	737	1 US-08-646-715-16	Sequence 16, Appl
4	216	12.2	921	1 US-08-188-582-2	Sequence 2, Appl
5	216	12.2	921	1 US-08-646-715-2	Sequence 2, Appl
6	176	9.9	2035	1 US-08-046-585-5	Sequence 5, Appl
7	176	9.9	2035	1 US-08-393-703-5	Sequence 5, Appl
8	176	9.9	2035	5 PCT-US93-11721-5	Sequence 5, Appl
9	164.5	9.3	732	1 US-08-244-189-2	Sequence 2, Appl
10	148.5	8.4	1721	3 US-08-700-651-5	Sequence 5, Appl
11	148.5	8.4	1721	3 US-08-928-3618-6	Sequence 5, Appl
12	144	8.1	941	4 US-07-757-0228-14	Sequence 14, Appl
13	144	8.1	1022	4 US-07-757-0228-74	Sequence 84, Appl
14	144	8.1	1038	4 US-07-757-0228-58	Sequence 58, Appl
15	144	8.1	1049	4 US-07-757-0228-46	Sequence 46, Appl
16	144	8.1	1140	4 US-07-757-0228-104	Sequence 104, Appl
17	144	8.1	1270	4 US-07-757-0228-44	Sequence 44, Appl
18	144	8.1	1311	4 US-07-757-0228-42	Sequence 42, Appl
19	144	8.1	1313	4 US-07-757-0228-142	Sequence 142, Appl
20	144	8.1	1314	4 US-07-757-0228-50	Sequence 50, Appl
21	144	8.1	1320	4 US-07-757-0228-46	Sequence 46, Appl
22	144	8.1	1320	4 US-07-757-0228-60	Sequence 60, Appl
23	144	8.1	1354	4 US-07-757-0228-48	Sequence 48, Appl
24	144	8.1	1361	4 US-07-757-0228-40	Sequence 40, Appl
25	144	8.1	1363	4 US-07-757-0228-52	Sequence 52, Appl
26	144	8.1	1404	4 US-07-757-0228-62	Sequence 62, Appl
27	144	8.1	1404	4 US-07-757-0228-62	Sequence 62, Appl

28	142.5	8.0	750	4 US-09-165-239A-4	Sequence 4, Appl
29	142.5	8.0	2476	2 US-08-276-967-2	Sequence 2, Appl
30	141	8.0	2972	4 US-09-579-181-2	Sequence 2, Appl
31	141	8.0	3118	4 US-09-579-181-1	Sequence 1, Appl
32	138.5	7.8	1837	3 US-08-928-3618-5	Sequence 5, Appl
33	137.5	7.8	907	3 US-08-783-774-2	Sequence 2, Appl
34	137.5	7.8	907	4 US-09-328-599A-1	Sequence 1, Appl
35	137.5	7.8	907	5 PCT-US95-04611A-19	Sequence 19, Appl
36	135.5	7.6	878	4 US-09-556-706B-2	Sequence 2, Appl
37	133	7.5	805	4 US-09-103-429A-4	Sequence 4, Appl
38	132.5	7.5	786	4 US-09-103-429A-3	Sequence 3, Appl
39	129	7.3	1481	2 US-08-616-844-40	Sequence 40, Appl
40	129	7.3	1481	2 US-08-599-654-40	Sequence 40, Appl
41	129	7.3	1481	3 US-08-944-688A-40	Sequence 40, Appl
42	129	7.3	1481	3 US-08-944-423A-40	Sequence 40, Appl
43	129	7.3	1481	3 US-08-944-496-40	Sequence 40, Appl
44	127.5	7.2	821	4 US-09-556-877-195	Sequence 195, App
45	127.5	7.2	821	4 US-09-620-412C-195	Sequence 195, App

ALIGNMENTS

```
RESULT 1
US-08-725-012-2
: Sequence 2, Application US/08725012
: Patent No. 5710025
: GENERAL INFORMATION:
: APPLICANT: Dikstein, Rivka
: APPLICANT: Tjian, Robert
: TITLE OF INVENTION: B-Cell Specific Transcription Factor
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
: STREET: 268 BUSH STREET, SUITE 3200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/725,012
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: OSMAN, RICHARD A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: B97-005
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 343-4341
: TELEFAX: (415) 343-4342
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 801 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-725-012-2

Query Match 100.0%; Score 1773; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 4.3e-155;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTLVTKVAPVAPKVSQGPLPAQIVAVKAPVTTTIOFPANQLPQGVLIKNSGPL 60
DB 1 GTLVTKVAPVAPKVSQGPLPAQIVAVKAPVTTTIOFPANQLPQGVLIKNSGPL 60
QY 61 MLVSPQQTVAETTSNTSPPAVPANDQVTKICVPSSSQLIKKAVTPVKKLAQIGT 120
```


ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-646-715-2

Query Match 12.2%; Score 216; DB 1; Length 921;
Best Local Similarity 24.9%; Pred. No. 2,6e-11;
Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;

QY 25 PQIVAVKAPNTTITQPPANLQLPPTVLIKNSGPMML-----VSPQOVTTRAETTS 76
DB 133 PQSPSTILSLNTNGQTPA-----LLVKTONGFOLLRVGTTGPTVTQTITNTSNN 184
QY 77 NITSRVAVNPQTVKICITPNSSQ-----LIKAVVPYKLAIGITVTTP 127
DB 185 NNTSTNHPPTTQ-IRLQVPAASMTNTTATSNITVNSVASSGYANSSOPHILTOLNAQ 243
QY 128 KPSSVAVAPTSVTVTPPKPLNTVTLTKPSLSGASTSPNEPNLKAENSAVQINLSP 187
DB 244 APOLPQITQITIPAOSSQOQVNNVSSAGTATVASTTA-----ATY 287
QY 188 TMLNPK-KCKNPLAMLIKACSGSOPENGQVKKLVEOLDLAKIEAEFTKRLYEKL 246
DB 288 TOGQNKKECKRKLAMLIETL-STREPKPVEKKNVRLIQLVNAVNEPEEFCRLLELLN 345
QY 247 SSPQPHLVPFLKKSVALRDL-----LNSQSFQ--- 276
DB 346 ASPOPLIGLKLSPRLKALYKELVIGICPPQHVIGLAGLSQQLKIDQIKRPIG 405
QY 277 ---OCVOQTSDMVATCTTVTTPVNTTVSSQSEKSIIVSGATAPTVS---VQT 329
DB 406 PSQTTTIGQTVNMI--TPNALGTPTTHTTISKQPN---IRLPTARLVNTGSI RT 460
QY 330 LNPAGPVGAKAGVTLHSVGP TAA 354
DB 461 QLP-SLQVPGQANIVQIR--GPQHA 482

RESULT 6
US-08-046-585-5
Sequence 5, Application US/08046585
Patent No. 5453362
GENERAL INFORMATION:
APPLICANT: Lamarco, Kelly
APPLICANT: Wilson, Angus
TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
TITLE OF INVENTION: HOST CELL FACTOR
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/046, 585
FILING DATE: 12-APR-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57503-1/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-046-585-5

Query Match 9.9%; Score 176; DB 1; Length 2035;
Best Local Similarity 24.2%; Pred. No. 4,3e-07;
Matches 108; Conservative 57; Mismatches 148; Indels 134; Gaps 22;

QY 10 VSAPPKVSGPRLPAPQIVAVKAPNTTITQPPANLQLPPTVLIKNSGPMMLVSPQOIV 69
DB 537 IGSSPQMSGMAALAAAAATOKIPSSA---PYLSVPACTTIVKT-----MAVTPTT 588
QY 70 TRAVTNTSRPVPANPOT-----VKICVPPNSSQ---LIKAVVPYKKL 115
DB 589 LPA--TVKAVSPVMSNPTRMLKTAADVOGVSVSATITSTRPITTVKSGTIVVAQ 646
QY 116 AQIGTIVV-----TVPRP-----SSVQSVAVPTS----- 140
DB 647 AQVTVTVGVGTITVLKPSISVGGGALISNLGVSVVQPKPVQTSVAVTQASTGCV 706
QY 141 -----VTVTPPKPLNTVTT-----LKPSSLGASST-----PSNE 170
DB 707 TQIIQTKPLPAGTILKLSADCKPTTITTTQOASGAGTKPITLIGISSVSPSTTKGTT 766
QY 171 PNKAENSAVQINLSPMLNPKKCNFLAML-IKLACSGSOPENGQVKKLVEOLD 229
DB 767 TIITKIMSALITQAGATGTSRGIKSPITITTKYMTSGTGP-----AKITVAV-- 818
QY 230 AKIEAEFTKRL-VYEKSSP-QP---HLVPLKKSVALROLPNSSQFIOQVOQTS 283
DB 819 PKIATGCGQGVTVLKGAPGCGTILRTVP-----MGVRLVTPVTVSAVPAV---- 869
QY 284 SDVATCTTIVTSPVNTTVSSQS-----EKSIIVSGATAPTVSVQTLNP 332
DB 870 -TTLVKGTTGVTVLGTGTGTVSTSLAGAGHSTASLSLATPITTLGTIA--TLLSOVINP 926
QY 333 LAGPVGAKAGVTLHSVGP TAA TGT 359
DB 927 TA--ITVSAQTLL-----TAAGGLTT 946

RESULT 7
US-08-393-703-5
Sequence 5, Application US/08393703
Patent No. 5585239
GENERAL INFORMATION:
APPLICANT: Lamarco, Kelly
APPLICANT: Wilson, Angus
TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
TITLE OF INVENTION: HOST CELL FACTOR
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA

```

: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/393,703
: FILING DATE: 24-FEB-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Osman, Richard A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: A-57503-2/RAO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2035 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-393-703-5

```

```

Query Match          9.9%; Score 176; DB 1; Length 2035;
Best Local Similarity 24.2%; Pred. No. 4.3e-07;
Matches 108; Conservative 57; Mismatches 148; Indels 134; Gaps 22;

```

```

QY 10 VSAPPKVSRRPAPQVAVKAPNTTIOFPANQLPFGVILKNSGPMILVSPQOTV 69
DB 537 IGSSPQKSGMAALAAAAATOKIPSSA---PTVLSPAGTTIVKT-----MAVPGTGT 588
QY 70 TRAEFTSNITSRPAPVAPNPQT-----VKICTVPSNSSQ---LIRKVAATPVKRL 115
DB 589 LPA--TVKVASPVVWNSPATRMLKTAAGVGTSSATNTSTRITIVHKSGLTVTAQO 646
QY 116 AQIGTTVV-----TVPKRP-----SSQVSAVPTS-----140
DB 647 AQVTVTVGVGVTKTITLVKSPISVPGSGALISNLKQVMSVQTKRVOGTSVATGASTGPV 706
QY 141 -----VVYTPGKPLNTVTT-----LKPSSIGAST-----PSNE 170
DB 707 TQIIQTKGPLAGTILIKVTSADGKPTIIITTTQASGAGTPTIIIGTSSVSPSTTKPGTT 766
QY 171 PNLKAENSAVOINLSPMLENVKCKNKLAML-IKLACSGSOSPQEMQNVKKLVEQLLD 229
DB 767 TIITIPMSALITQAGATGVTSSPGIKSPIITITTKVMTSGTAGP-----AKIITAV-- 818
QY 230 AKIAEEFTRL-YVELKSSP-QP-----HLVPFLKSVVALROLNPSQSFIOQCVOQTS 283
DB 819 PKIATGHQOGCVTVLKGAGQPGTILRTVP-----MGVRLVTPVVSAAVKPAAV----- 869
QY 284 SDMYATCTTIVTSPVTTIVSSSQS-----EKSIIVSGATAPRTVSQIOLNP 332
DB 870 -TTLVVKGTGTGTVTLGTVTSTSLAGAGHSTASLAPITTLGLTIA--TLSSQVYNP 926
QY 333 LAGPVGAKAGVTLHSVGPATAGTGT 359
DB 927 TA--ITVSAQOTL-----TAAAGLIT 946

```

```

RESULT 8
PCT-US93-11721-5
: Sequence 5, Application PC/TUS9311721
: GENERAL INFORMATION:
: APPLICANT: Lamarco, Kelly
: APPLICANT: Wilson, Angus
: APPLICANT: Herr, Winship
: TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:

```

```

: TITLE OF INVENTION: HOST CELL FACTOR
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/11721
: FILING DATE: 03-DEC-1993
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Osman, Richard A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: PP-57503-1/RAO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2035 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: PCT-US93-11721-5

```

```

Query Match          9.9%; Score 176; DB 5; Length 2035;
Best Local Similarity 24.2%; Pred. No. 4.3e-07;
Matches 108; Conservative 57; Mismatches 148; Indels 134; Gaps 22;

```

```

QY 10 VSAPPKVSRRPAPQVAVKAPNTTIOFPANQLPFGVILKNSGPMILVSPQOTV 69
DB 537 IGSSPQKSGMAALAAAAATOKIPSSA---PTVLSPAGTTIVKT-----MAVPGTGT 588
QY 70 TRAEFTSNITSRPAPVAPNPQT-----VKICTVPSNSSQ---LIRKVAATPVKRL 115
DB 589 LPA--TVKVASPVVWNSPATRMLKTAAGVGTSSATNTSTRITIVHKSGLTVTAQO 646
QY 116 AQIGTTVV-----TVPKRP-----SSQVSAVPTS-----140
DB 647 AQVTVTVGVGVTKTITLVKSPISVPGSGALISNLKQVMSVQTKRVOGTSVATGASTGPV 706
QY 141 -----VVYTPGKPLNTVTT-----LKPSSIGAST-----PSNE 170
DB 707 TQIIQTKGPLAGTILIKVTSADGKPTIIITTTQASGAGTPTIIIGTSSVSPSTTKPGTT 766
QY 171 PNLKAENSAVOINLSPMLENVKCKNKLAML-IKLACSGSOSPQEMQNVKKLVEQLLD 229
DB 767 TIITIPMSALITQAGATGVTSSPGIKSPIITITTKVMTSGTAGP-----AKIITAV-- 818
QY 230 AKIAEEFTRL-YVELKSSP-QP-----HLVPFLKSVVALROLNPSQSFIOQCVOQTS 283
DB 819 PKIATGHQOGCVTVLKGAGQPGTILRTVP-----MGVRLVTPVVSAAVKPAAV----- 869
QY 284 SDMYATCTTIVTSPVTTIVSSSQS-----EKSIIVSGATAPRTVSQIOLNP 332
DB 870 -TTLVVKGTGTGTVTLGTVTSTSLAGAGHSTASLAPITTLGLTIA--TLSSQVYNP 926
QY 333 LAGPVGAKAGVTLHSVGPATAGTGT 359
DB 927 TA--ITVSAQOTL-----TAAAGLIT 946

```

```

RESULT 9

```


COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/928,361B
 FILING DATE: 12-SEP-1997
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/026,062
 FILING DATE: 13-SEP-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Vervy, Hana
 REGISTRATION NUMBER: 30,518
 REFERENCE/DOCKET NUMBER: 480.76-1(HV)
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-324-1677
 TELEFAX: 650-324-1678
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1721 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-928-361B-6

Query Match 8.4%; Score 148.5; DB 3; Length 1721;
 Best Local Similarity 21.9%; Pred. No. 0.00012;
 Matches 82; Conservative 41; Mismatches 183; Indels 69; Gaps 12;

QY 35 TTITQFANLQLPPEYVLIKNSGPMMLVSPQOTVRAETTSNITSRAVPANPQTVKIC 94
 DB 286 TTTTITT 345
 QY 95 TVPRSSQLIKKVAATPVKLAQIGTVTVTPKPSVQSAVPTSVYVTPGKPLMTVT 154
 DB 346 TTTTITT 398
 QY 155 TLKPSISGASTPS-----NEPNLKAENSA-----VOINLSPMLEN-- 192
 DB 399 TTTATTTTTTSETESVIRPDEMCKEKNNGEAKATYGVYIGKGRLENQAFMIRDD 458
 QY 193 -----VKCKNFLAMLIKACSGSOSPEMGQNVKKLVEQLDARIE--AEETRLKY 242
 DB 459 THVFERKVDVGNITISVRCNKGAKLEFP-----DRSLDFTLPVAGHNSCII 508
 QY 243 VELKSSQPHLVPFLKSYVALRQLPNSQSF-----IQOCY-----QOTSSDWIA 289
 DB 509 VGVSGDKIHVSPYSGKDVLSIAPISPELFENEVCDDCTAKYGAIHSGYQTSADPVT 568
 QY 290 TC-----TTTTTSPVYTTTVSSQSEKSIIVSGATAPRTVSQVTLNPLAGPVGAKGVVTL 346
 DB 569 TTKAPTTTTTGABQGPPTTTTSGPSKP--TTTTTKATTTTTTILPITTTTQKPTTTT 626
 QY 347 HSV--GPTAATGTT 359
 DB 627 TKVGGKPIATTTTT 641

RESULT 12
 US-07-757-022B-14
 Sequence 14, Application US/07757022B
 Patent No. 6431142
 GENERAL INFORMATION:

APPLICANT: Gesner, Thomas G.
 APPLICANT: Clark, Stephen C.
 APPLICANT: Turner, Katherine
 APPLICANT: Hewick, Rodney M.
 TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 Cambridgepark Drive

CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/757,022B
 FILING DATE: 19910910
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Cseri, Luanu
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: G1 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)876-1170
 TELEFAX: (617)876-5851
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 941 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-757-022B-14

Query Match 8.1%; Score 144; DB 4; Length 941;
 Best Local Similarity 20.3%; Pred. No. 0.00012;
 Matches 75; Conservative 53; Mismatches 159; Indels 82; Gaps 12;

QY 8 APVSAAPRVSSQPHLPAQIVAVKAPNT-----TTIOPANLQLPPEYVLI 53
 DB 482 APTTPEKAPPTTPEKAPPTTPEKAPPTTPEKAPPTTPEKAPPTTPEKAPPTT 538
 QY 54 KNSGPMMLVSPQOTVRAETTSNITSRAVPANPQTVKICTVPSQSLIKKVAATPVK 113
 DB 539 --TKEPSTSDKRAPPTPKGTAPPTTPEKAPPTTPEKAPPTTPEKAPPTTPEKAPPT 596
 QY 114 KLA-----QIGTVTVTVTPKPSVQSAV--VPTSVTVTPGKPLMTVTVTLKPSLG 162
 DB 597 KPAKELAPTTTGTSTSDKPAPTTPEKAPPTTPEKAPPTTPEKAPPTTPEKAPPT 656
 QY 163 ASSTP--SNEPNL--KNSAAYOINLSP--MLEN-----VKCKNFLAMLIKACSG 210
 DB 657 EVSTPTTKEPTTTHKSDESTPELSAEPYKALENSKEPVGPTTTPAATKPEMTT 716
 QY 211 -----SOSPEMGQNVKKLVEQLDARIEAEETRLKYVELKSSQPHLV 254
 DB 717 KDKTTERDLRTTPEPTTAPAKMTKETATTEKTESKITAT-----TQVTSITTOOT 770
 QY 255 PFLKSYVALRQLPNSQSFIOOCVQOTSDMVATCTTTVTSPVYTTVSSQSEKSI 314
 DB 771 PF-KITTLKTYTLAPK-----VTTTKKITITTEIMNKKPEETAKPKDRA 812
 QY 315 IVSGATAPR 323
 DB 813 TNSKATTPK 821

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 22:01:05 ; Search time 38.818 Seconds
(without alignments)
1827.005 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_552

Perfect score: 562
Sequence: 1 AQPpVLSPAGIPGTGSSSK.....ASPTOKNRKIKENTSCFRDE 110

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 4569144 segs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending_Patents_AA_Main: *
1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep: *
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep: *
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep: *
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep: *
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep: *
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep: *
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep: *
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep: *
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep: *
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep: *
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep: *
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep: *
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep: *
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep: *
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep: *
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep: *
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep: *
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep: *
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep: *
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep: *
21: /cgn2_6/ptodata/2/paa/US097_COMB.pep: *
22: /cgn2_6/ptodata/2/paa/US098_COMB.pep: *
23: /cgn2_6/ptodata/2/paa/US099_COMB.pep: *
24: /cgn2_6/ptodata/2/paa/US100_COMB.pep: *
25: /cgn2_6/ptodata/2/paa/US101_COMB.pep: *
26: /cgn2_6/ptodata/2/paa/US102_COMB.pep: *
27: /cgn2_6/ptodata/2/paa/US60_COMB.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	562	100.0	801	1	PCT-US01-08631-39827
2	562	100.0	852	21	US-09-763-909-2
3	534	95.0	843	1	PCT-US02-29864-410
4	493.5	87.8	865	1	PCT-US02-25829-25
5	302	53.7	685	27	US-60-243-468-1271
6	92.5	16.5	737	1	PCT-US94-01114-16

	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	
	92.5	92.5	92.5	84.5	84.5	80	80	80	78.5	78.5	78.5	78.5	78.5	78.5	78.5	77.5	77.5	77.5	77.5	76.5	76.5	76.5	76.5	75.5	75.5	75.5	74.5	74.5	74.5	74.5	73.5	73.5	73.5	73.5	73.5	73.5	73.5	73.5	73.5	73.5
	16.5	16.5	16.5	15.0	15.0	14.2	14.2	14.2	14.0	14.0	14.0	14.0	14.0	14.0	14.0	13.8	13.8	13.8	13.8	13.6	13.6	13.6	13.6	13.5	13.5	13.5	13.4	13.4	13.4	13.3	13.3	13.1	13.1	13.0	13.0	13.0	13.0	13.0	13.0	
	1051	1051	1051	1083	1083	624	624	624	471	471	471	471	471	471	471	3726	3726	3726	3726	3726	3726	3726	3726	3065	3065	3065	3085	3085	3085	3085	3085	3085	3085	3085	3085	3085	3085	3085	3085	
	1	1	1	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	PCT-US02-30474-1660	US-60-324-631-1665	PCT-US02-30474-1665	US-60-389-987-1773	US-60-412-418-1773	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	US-09-248-796-14351	
	Sequence 1660, Ap	Sequence 1665, Ap	Sequence 3271, Ap	Sequence 1773, Ap	Sequence 1773, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	Sequence 14351, Ap	

ALIGNMENTS

RESULT 1
PCT-US01-08631-39827
Sequence 39827, Application PC/TUS0108631
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc
FILE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
TITLE REFERENCE: 21272-049
CURRENT APPLICATION NUMBER: PCT/US01/08631
CURRENT FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: 09/540,217
PRIOR FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: 09/649,167
PRIOR FILING DATE: 2000-08-23
NUMBER OF SEQ ID NOS: 60736
SOFTWARE: Custom
SEQ ID NO 39827
LENGTH: 801
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US01-08631-39827

Query Match 100.0%; Score 562; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 7.6e-53;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AQPpVLSPAGIPGTGSSSKQLFLFHVQOPSGNGKQVTTISHSSTLTITOKGCGATMP 60
DB 443 AQPpVLSPAGIPGTGSSSKQLFLFHVQOPSGNGKQVTTISHSSTLTITOKGCGATMP 502

QY 61 VNTIIPSOFPASILKQITLPGNKILSLQASPTQKNRIKENVTSQFRDE 110
DB 503 VNTIIPSOFPASILKQITLPGNKILSLQASPTQKNRIKENVTSQFRDE 552

RESULT 2

US-09-763-909-2
Sequence 2, Application US/09763909
GENERAL INFORMATION:
APPLICANT: Dikstein, Rivka
TITLE OF INVENTION: A TRANSCRIPTION FACTOR TFIID SUBUNIT,
TITLE OF INVENTION: TAFII105, POLYPEPTIDES, DNA ENCODING THEREFOR AND
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS
FILE REFERENCE: 13005/002001
CURRENT APPLICATION NUMBER: US/09/763,909
CURRENT FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 852
TYPE: PRT
ORGANISM: Homo sapiens
US-09-763-909-2

Query Match 100.0%; Score 562; DB 21; Length 852;
Best Local Similarity 100.0%; Pred. No. 8,3e-53;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AOPGPVLSOPAGIPTSSSKQLFSLFHVVOOPSGNGEKQVTTSHSSTLTIOKCGQKTP 60
DB 443 AOPGPVLSOPAGIPTSSSKQLFSLFHVVOOPSGNGEKQVTTSHSSTLTIOKCGQKTP 502

QY 61 VNTIIPSOFPASILKQITLPGNKILSLQASPTQKNRIKENVTSQFRDE 110
DB 503 VNTIIPSOFPASILKQITLPGNKILSLQASPTQKNRIKENVTSQFRDE 552

RESULT 3

PCT-US02-29964-410
Sequence 410, Application PC/TUS0229964
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Ren, Feiyun
APPLICANT: Zhang, Jie
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Jian-Rui
APPLICANT: Xue, Aidong J.
APPLICANT: Wang, Duncui
APPLICANT: Ghosh, Malabika
APPLICANT: Asundi, Vinod
APPLICANT: Wehrman, Tom
APPLICANT: Zhou, Ping
APPLICANT: Goodrich, Ryle W.
APPLICANT: Meng, Gezhi
APPLICANT: Haley-Vicente, Dana
APPLICANT: Drmanac, Radote T
TITLE OF INVENTION: Novel Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 809ACIP PCT
CURRENT APPLICATION NUMBER: PCT/US02/29964
CURRENT FILING DATE: 2002-09-19
PRIOR APPLICATION NUMBER: US 60/323,739
PRIOR FILING DATE: 2001-09-19
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02623

PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: US 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: PCT/US01/03800
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 09/515,126
PRIOR FILING DATE: 2000-02-28
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 992
SOFTWARE: PL_FL_genes Version 6.0
SEQ ID NO 410
LENGTH: 843
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US02-29964-410

Query Match 95.0%; Score 534; DB 1; Length 843;
Best Local Similarity 95.5%; Pred. No. 1.1e-49;
Matches 105; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AOPGPVLSOPAGIPTSSSKQLFSLFHVVOOPSGNGEKQVTTSHSSTLTIOKCGQKTP 60
DB 485 AOPGPVLSOPAGIPQAVQKQLFSLFHVVOOPSGNGEKQVTTSHSSTLTIOKCGQKTP 544

QY 61 VNTIIPSOFPASILKQITLPGNKILSLQASPTQKNRIKENVTSQFRDE 110
DB 545 VNTIIPSOFPASILKQITLPGNKILSLQASPTQKNRIKENVTSQFRDE 594

RESULT 4

PCT-US02-25829-25
Sequence 25, Application PC/TUS0225829
GENERAL INFORMATION:
APPLICANT: INCYTE GENOMICS, INC.
APPLICANT: BAROSSO, Ines
APPLICANT: BAUGHN, Mariah R.
APPLICANT: BECHA, Shanya D.
APPLICANT: BLAKE, Julie J.
APPLICANT: BOROMSKY, Mark L.
APPLICANT: BUREFORD, Neil
APPLICANT: DUGGAN, Brendan M.
APPLICANT: ELLIOTT, Vicki S.
APPLICANT: EMERLING, Brooke M.
APPLICANT: FORSYTHE, Ian J.
APPLICANT: GIETZEN, Kimberly J.
APPLICANT: GORVAD, Ann E.
APPLICANT: GRIFFIN, Jennifer A.
APPLICANT: HAFALIA, April J.A.
APPLICANT: HONCHELL, Cynthia D.
APPLICANT: ISON, Craig H.
APPLICANT: KHAN, Farrah A.
APPLICANT: LAL, Preeti G.
APPLICANT: LEE, Ernestine A.
APPLICANT: LEE, Sally
APPLICANT: LEE, Soo Yeun
APPLICANT: LI, Joana X.
APPLICANT: LU, Dzung Alina M.
APPLICANT: LU, Yan
APPLICANT: LEHR-MASON, Patricia M.
APPLICANT: NGUYEN, Daniel B.
APPLICANT: RAMKUMAR, Jayalaxmi
APPLICANT: SPRAGUE, William W.
APPLICANT: TANG, Y. Tom
APPLICANT: THANGAVELU, Kavitha
APPLICANT: THORNTON, Michael
APPLICANT: TRAN, Uyen K.
APPLICANT: WALIA, Narinder K.
APPLICANT: WARREN, Bridget A.
APPLICANT: XU, Yuning
APPLICANT: YAO, Montique G.
APPLICANT: YUE, Henry

APPLICANT: YUE, Huibin
APPLICANT: ZEBARJADIAN, Yeganeh
TITLE OF INVENTION: NUCLEIC ACID ASSOCIATED PROTEINS
FILE REFERENCE: PF-1146 PCT
CURRENT APPLICATION NUMBER: PCT/US02/25829
CURRENT FILING DATE: 2002-08-14
PRIOR APPLICATION NUMBER: US 60/313,111
PRIOR FILING DATE: 2001-08-17
PRIOR APPLICATION NUMBER: US 60/314,682
PRIOR FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/314,756
PRIOR FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/315,105
PRIOR FILING DATE: 2001-08-27
PRIOR APPLICATION NUMBER: US 60/316,751
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: US 60/316,856
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: US 60/328,185
PRIOR FILING DATE: 2001-10-05
NUMBER OF SEQ ID NOS: 66
SOFTWARE: PERL Program
SEQ ID NO 25
LENGTH: 865
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: 4398735CD1
PCT-US02-25829-25

Query Match 87.8%; Score 493.5; DB 1; Length 865;
Best Local Similarity 90.9%; Pred. No. 3.6e-45;
Matches 100; Conservative 0; Mismatches 5; Indels 5; Gaps 1;

OY 1 AOPGVLSQAPGIPPTGSSSKOLFSLFHVQOPSGGNEKOVTTISHSTLTIOKCGOKTNP 60
|||||
DB 512 AOPGVLSQAPGIPQAVGVKOL-----VVOOPSGGNEKOVTTISHSTLTIOKCGOKTNP 566
|||||
OY 61 VNTIIPTSQPPASILKQITLPGNKILSLQASPTQKNRIKENVTSCEFDE 110
|||||
DB 567 VNTIIPTSQPPASILKQITLPGNKILSLQASPTQKNRIKENVTSCEFDE 616
|||||

RESULT 5
US-60-243-468-1271
Sequence 1271, Application US/60243468
GENERAL INFORMATION:
APPLICANT: Beasley, Ellen
TITLE OF INVENTION: ISOLATED HUMAN DRUG TARGET PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN DRUG TARGET PROTEINS,
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: CL000929
CURRENT APPLICATION NUMBER: US/60/243,468
CURRENT FILING DATE: 2000-10-27
NUMBER OF SEQ ID NOS: 2121
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1271
LENGTH: 685
TYPE: PRT
ORGANISM: HUMAN
US-60-243-468-1271

Query Match 53.7%; Score 302; DB 27; Length 685;
Best Local Similarity 68.8%; Pred. No. 5.4e-24;
Matches 66; Conservative 6; Mismatches 18; Indels 6; Gaps 2;

OY 1 AOPGVLSQAPGIPPTGSSSKOLFSLFHVQOPSGGNEKOVTTISHSTLTIOKCGOKTNP 60
|||||
DB 473 AOPGVLSQAPGIPQAVGVKOL-----VVOOPSGGSEKOVTTISHSTLTIOKCGOKTNP 527
|||||
OY 61 VNTIIPTSQPPASILKQITLPGNKILSLQASPTQKNRIKENVTSCEFDE 96
|||||

DB 528 VNTIIPTSQPPGK-KHDITELNSDAVNLISQATOE 562
PCT-US94-01114-16
Sequence 16, Application PC/TUS9401114
GENERAL INFORMATION:
APPLICANT: Tian, Robert
APPLICANT: Comal, Lucio
APPLICANT: Dynlacht Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Rupert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TATS AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01114
FILING DATE: 28-JAN-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: FP-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 737 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01114-16

Query Match 16.5%; Score 92.5; DB 1; Length 737;
Best Local Similarity 27.3%; Pred. No. 1.3; Length 737;
Matches 33; Conservative 19; Mismatches 50; Indels 19; Gaps 5;

OY 2 OPGVLS--QAPGIPPTGSSSKOLFSLFHVQOPSGGNEKOVTTISHSTLTIOKCGOKTNP 59
|||||
DB 370 QP-PVLSLQTPQVGVKQGPPLVIOGPKPKGALINPQVTLQTFVVALRQ----- 422
|||||
OY 60 PVNTIIP-----SQPPASILKQITLPGNKIL--SLQASPTQKNRIKENVTSCEFD 109
|||||
DB 423 PHNRIMLTTPQIQINLPQVVPVAVLPGRKALSAVSAQAAMAKNKLKPGGSGSPFD 482
|||||
OY 110 E 110
DB 483 D 483

RESULT 7
PCT-US02-30474-1660
Sequence 1660, Application PC/TUS0230474
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Zhang, Jie

```
APPLICANT: Ren, Feiyan
APPLICANT: xue, Aidong J.
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Jian-Rui
APPLICANT: Wehrman, Tom
APPLICANT: Zhou, Ping
APPLICANT: Ghosh, Malabika
APPLICANT: Wang, Dunrui
APPLICANT: Ma, Yundong
APPLICANT: Asundi, Vinod
APPLICANT: Wang, Zhiwei
APPLICANT: Weng, Gezhi
APPLICANT: Haley-Vicente, Dana
APPLICANT: Drmanac, Radoje T
TITLE OF INVENTION: Novel Nucleic Acids and
FILE REFERENCE: Polypeptides
CURRENT FILING DATE: 2001-09-24
PRIOR APPLICATION NUMBER: US 60/324,631
PRIOR FILING DATE: 2001-09-24
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02623
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: US 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: PCT/US01/03800
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 09/515,126
PRIOR FILING DATE: 2000-02-28
Remainder prior Application data removed - See file wrapper or PALM.
NUMBER OF SEQ ID NOS: 3476
SOFTWARE: PL_FL_genes Version 6.0
SEQ ID NO 1660
LENGTH: 1051
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US02-30474-1660

Query Match      16.5%; Score 92.5; DB 1; Length 1051;
Best Local Similarity 27.3%; Pred. No. 2.1;
Matches 33; Conservative 19; Mismatches 50; Indels 19; Gaps 5;

QY 2 QGPGVLS--QPGAGPTGSSSKQLFSEHVVOQPSGNGNEKQVTTISHSSTLTIKCGQKTM 59
DB 684 QP-PVLSLQGPVGVGKQGPPLVLIQPPKRGALIRPPQVTLTQTPMALNQ----- 736
QY 60 PVTITPT-----SQFPASILKQITLPGNKIL--SLQASPTQNRKIKENTSCFRD 109
DB 737 PHRRIMLTTPQOIQILNPLQPVVPRKPAVLPGRKALSAVSAQAAAKNKLKKEGGGSRD 796
QY 110 E 110
DB 797 D 797

RESULT 8
US-60-324-631-1665
; Sequence 1665, Application US/60324631
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Xue, Aidong J.
```

```
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Jian-Rui
APPLICANT: Wehrman, Tom
APPLICANT: Zhou, Ping
APPLICANT: Ghosh, Malabika
APPLICANT: Wang, Dunrui
APPLICANT: Ma, Yundong
APPLICANT: Asundi, Vinod
APPLICANT: Wang, Zhiwei
APPLICANT: Weng, Gezhi
APPLICANT: Haley-Vicente, Dana
APPLICANT: Drmanac, Radoje T
TITLE OF INVENTION: Novel Nucleic Acids and
FILE REFERENCE: Polypeptides
CURRENT FILING DATE: 2001-09-24
PRIOR APPLICATION NUMBER: US 60/324,631
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02623
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: US 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: PCT/US01/03800
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 09/515,126
PRIOR FILING DATE: 2000-02-28
PRIOR APPLICATION NUMBER: US 09/577,409
PRIOR FILING DATE: 2000-05-18
PRIOR APPLICATION NUMBER: PCT/US01/04927
PRIOR FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: US 09/519,705
PRIOR FILING DATE: 2000-03-07
PRIOR APPLICATION NUMBER: US 09/574,454
PRIOR FILING DATE: 2000-05-19
PRIOR APPLICATION NUMBER: PCT/US01/04941
PRIOR FILING DATE: 2001-03-05
PRIOR APPLICATION NUMBER: US 09/540,217
PRIOR FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: US 09/649,167
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: PCT/US01/08631
PRIOR FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: US 09/552,929
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: US 09/770,160
PRIOR FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: PCT/US01/08656
PRIOR FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: US 09/577,408
PRIOR FILING DATE: 2000-05-18
PRIOR APPLICATION NUMBER: PCT/US01/14827
PRIOR FILING DATE: 2001-05-16
NUMBER OF SEQ ID NOS: 3334
SOFTWARE: PL_FL_genes Version 6.0
SEQ ID NO 1665
LENGTH: 1051
TYPE: PRT
ORGANISM: Homo sapiens
US-60-324-631-1665

Query Match      16.5%; Score 92.5; DB 27; Length 1051;
Best Local Similarity 27.3%; Pred. No. 2.1;
Matches 33; Conservative 19; Mismatches 50; Indels 19; Gaps 5;
```

```

QY      2  QPGVSL--CPAGIPGTSSSKQLFSLPHVYVQPPGGGNEKQVTTISHSTLTIOKCGKTM 59
      || ||| || | : : : : :
Db      684  QP-FVLSLTQPTQVGVGKQGPPLVLIQDPKKPQALIRPQVTTLTQTPMALNQ----- 736
QY      60  PVNTIIP-----SOPPPASILKQITLPGNKIL--SLQASPTOKRIKENVTSQCFRD 109
      | | : : : | ||| | | | : ||| | |
Db      737  PHNMIMLTTPQIOLNPLQPIPVYKPAVLFTGTRLSAVSQAAAAQKNKKLKEPGGGSFRD 796
QY      110  E 110
      :
Db      797  D 797

```

```

RESULT 9
PCT-US02-30474-3271
Sequence 3271, Application PC/TUS0230474
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Xue, Aйдong J.
APPLICANT: Zhao, Qiong A.
APPLICANT: Wang, Jіan-Rui
APPLICANT: Weinman, Tom
APPLICANT: Zhou, Ping
APPLICANT: Ghosh, Malabika
APPLICANT: Wang, Dunrui
APPLICANT: Ma, Yundong
APPLICANT: Asundi, Vinod
APPLICANT: Wang, Zhilwei
APPLICANT: Wang, Gezhi
APPLICANT: Haley-Vicente, Dana
APPLICANT: Drmanac, Radoje T
TITLE OF INVENTION: Novel Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 810CIP PCT
CURRENT APPLICATION NUMBER: PCT/US02/30474
CURRENT FILING DATE: 2002-09-24
PRIOR APPLICATION NUMBER: US 60/324,631
PRIOR FILING DATE: 2001-09-24
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02623
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: US 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: PCT/US01/03800
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 09/515,126
PRIOR FILING DATE: 2000-02-28
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 3476
SOFTWARE: pl_fl_genes Version 6.0
SEQ ID NO 3271
LENGTH: 1083
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(1083)
OTHER INFORMATION: Xaa = any amino acid or symbol as shown in the table 8 as set forth
PCT-US02-30474-3271

Query Match 16.5%; Score 92.5; DB 1; Length 1083;
Best Local Similarity 27.3%; Pred. No. 2.2;

```

Matches	33, Conservative	19, Mismatches	50, Indels	19, Gaps	5, ...
QY	2	QPGPVLS--QPGAGIPGSSSKQLSFLFHVVOOPSGGNEKQVTTISHSSTLTITQKCGQKTM	59		
Db	716	QP-PVLSTLTQPTQVGVGQGOPTPLVIOQPPPGALIRPPQVTLTQPTVMVALRQ-----	768		
QY	60	PVNTIIP-----SQFPASILKQITLPGKIL---SLQSPQKNRIKENYVSCFPD	109		
Db	769	PHNRMLTTPPOQIQLNPVQPVVAVKPAVLPGTKALSAVSAQAAAAQKNKLEKPGGSGSPD	828		
QY	110	E 110			
Db	829	D 829			

```

RESULT 10
US-60-389-987-1773
Sequence 1773, Application US/60389987
GENERAL INFORMATION:
APPLICANT: Ghosh, Soumitra S.
APPLICANT: Fahy, Eoin D.
APPLICANT: Zhang, Bing
APPLICANT: Gibson, Bradford W.
APPLICANT: Taylor, Steven W.
APPLICANT: Glenn, Gary W.
APPLICANT: Warnock, Dale E.
TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
FILE REFERENCE: 660088.465P2
CURRENT APPLICATION NUMBER: US/60/389,987
CURRENT FILING DATE: 2002-06-17
NUMBER OF SEQ ID NOS: 3025
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 1773
LENGTH: 624
TYPE: PRT
ORGANISM: Homo sapiens
US-60-389-987-1773

Query Match          15.0%; Score 84.5; DB 27; Length 624;
Best Local Similarity 28.9%; Pred. No.7.9;
Matches 37; Conservative 11; Mismatches 41; Indels 39; Gaps 7;

QY      3  GGVLPASGAGIP-----TGSSSKQLSLFHHVQPSGGENKQVTTISHSTLTIOKCGOK 57
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      302  GGPASCGGGRQPLPSATGDR-----RGVGPQSWGPEAQAASASSPLEALAC-LK 352

QY      58  TNPVNTTTPTSQFPFASILKQITLPGKNKILSLQASPTOKNRKEN----- 102
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      353  GIPPGSSP-SQLPPTS-CSQNPQPGD-----SRQKPELOPHRSHSEATREPVLP 403

QY      103  -VTSQFRD 109
      : || ||
Db      404  GIGSCVRD 411

RESULT 11
US-60-412-418-1773
Sequence 1773, Application US/60412418
GENERAL INFORMATION:
APPLICANT: Ghosh, Soumitra S.
APPLICANT: Fahy, Eoin D.
APPLICANT: Zhang, Bing
APPLICANT: Gibson, Bradford W.
APPLICANT: Taylor, Steven W.
APPLICANT: Glenn, Gary W.
APPLICANT: Warnock, Dale E.
TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
FILE REFERENCE: 660088.465P3
CURRENT APPLICATION NUMBER: US/60/412,418
CURRENT FILING DATE: 2002-09-20
NUMBER OF SEQ ID NOS: 3025

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 22:00:19 ; Search time 4.47537 Seconds
(without alignments)
2076.603 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_552

Perfect score: 562
Sequence: 1 AAGPVLSPAGIPGSSSK.....ASPTOKRIKENVTSCEFRDE 110

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 429898 seqs, 84487048 residues

Total number of hits satisfying chosen parameters: 429898

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	18.0	925	US-09-724-676-63928	Sequence 63928, A
2	101	18.0	925	US-09-724-676-63928	Sequence 63928, A
3	101	18.0	957	US-09-724-676-63931	Sequence 63931, A
4	101	18.0	957	US-09-724-676-63931	Sequence 63931, A
5	101	18.0	1021	US-09-724-676-63922	Sequence 63922, A
6	101	18.0	1021	US-09-724-676-63922	Sequence 63922, A
7	101	18.0	1053	US-09-724-676-63925	Sequence 63925, A
8	101	18.0	1053	US-09-724-676-63925	Sequence 63925, A
9	93	16.5	679	US-09-724-676-63929	Sequence 63929, A
10	93	16.5	679	US-09-724-676-63929	Sequence 63929, A
11	93	16.5	711	US-09-724-676-63920	Sequence 63920, A
12	93	16.5	711	US-09-724-676-63920	Sequence 63920, A
13	93	16.5	775	US-09-724-676-63923	Sequence 63923, A
14	93	16.5	775	US-09-724-676-63923	Sequence 63923, A
15	93	16.5	807	US-09-724-676-63926	Sequence 63926, A
16	93	16.5	807	US-09-724-676-63926	Sequence 63926, A
17	84.5	15.0	616	US-10-245-882-264	Sequence 264, App
18	84.5	15.0	616	US-10-245-882-264	Sequence 264, App
19	84.5	15.0	709	US-09-724-676-63927	Sequence 63927, A
20	84.5	15.0	709	US-09-724-676-63927	Sequence 63927, A
21	84.5	15.0	741	US-09-724-676-63930	Sequence 63930, A
22	84.5	15.0	741	US-09-724-676-63930	Sequence 63930, A
23	84.5	15.0	805	US-09-724-676-63921	Sequence 63921, A
24	84.5	15.0	805	US-09-724-676-63921	Sequence 63921, A
25	84.5	15.0	837	US-09-724-676-63924	Sequence 63924, A
26	84.5	15.0	837	US-09-724-676-63924	Sequence 63924, A

27	84.5	15.0	1030	1	PCT-US02-29560-263	Sequence 263, App
28	84.5	15.0	1030	6	US-10-245-882-263	Sequence 263, App
29	75.5	13.4	441	5	US-09-134-000C-6703	Sequence 6703, App
30	75.5	13.4	441	5	US-09-134-000C-6703	Sequence 6703, App
31	75.5	13.4	557	5	US-09-134-000C-5061	Sequence 5061, App
32	75.5	13.4	557	5	US-09-134-000C-5061	Sequence 5061, App
33	74	13.2	617	6	US-10-144-779-555	Sequence 555, App
34	72.5	12.9	412	6	US-10-092-411A-3773	Sequence 3773, App
35	71	12.6	528	1	PCT-US02-04915-14	Sequence 14, App
36	70.5	12.5	128	1	PCT-US02-33727-7803	Sequence 7803, App
37	70.5	12.5	128	6	US-10-057-498-7803	Sequence 7803, App
38	70	12.5	410	5	US-09-751-708A-48	Sequence 48, App
39	69.5	12.4	1532	6	US-10-218-140-1418	Sequence 1418, App
40	68.5	12.2	186	5	US-09-724-676-87265	Sequence 87265, A
41	68.5	12.2	186	5	US-09-724-676-87265	Sequence 87265, A
42	68.5	12.2	213	5	US-09-724-676-87264	Sequence 87264, A
43	68.5	12.2	213	5	US-09-724-676-87264	Sequence 87264, A
44	68.5	12.2	579	5	US-09-724-676-87260	Sequence 87260, A
45	68.5	12.2	579	5	US-09-724-676-87260	Sequence 87260, A

ALIGNMENTS

RESULT 1
US-09-724-676-63928
; Sequence 63928, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: CompuGen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 CompuGen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63928
; LENGTH: 925
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-63928

Query Match 18.0%; Score 101; DB 5; Length 925;
Best Local Similarity 29.2%; Pred. No. 0.14;
Matches 38; Conservative 16; Mismatches 46; Indels 30; Gaps 5;

QY 2 QGPGVLSQAPAGIPG-----SSSKQLFSLFHYQDSGGEKRYT-----TISHSSRLT 50
DB 551 QPPPTSOATTALTAVALVSSS-----VQRTAGKTAATVTSALQPPVLSLTQVQVG 601
QY 51 IQKCGOKTMY-----NIIPTSQFPASILKQITLPGNKIL---SLQASPTOKNRK 100
DB 602 VGKGGQPRPLVYIQPPKRGALIQNLPQPPVPAVPLPGTKALSAVSAQAAAAQKNTK 661
QY 101 ENVTSCFRDE 110
DB 662 EPGGGSFRDD 671

RESULT 2
US-09-724-676A-63928
; Sequence 63928, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: CompuGen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 CompuGen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63928
; LENGTH: 925
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-724-676A-63928

Query Match	18.0%;	Score 101;	DB 5;	Length 925;
Best Local Similarity	29.28;	Pred. No. 0.14;		
Matches 38;	Conservative 16;	Mismatches 46;	Indels 30;	Gaps 5

[illegible]

RESULT 3

```

: Sequence 63931, Application US/09724676
: GENERAL INFORMATION:
:   APPLICANT: Compugen LTD
:   TITLE OF INVENTION: Variants of alternative splicing
:   FILE REFERENCE: 129181.4 Compugen
:   CURRENT APPLICATION NUMBER: US/09/724,676
:   CURRENT FILING DATE: 2000-11-28
:   NUMBER OF SEQ ID NOS: 97222
:   SOFTWARE: PatentIn version 3.2
:   SEQ ID NO 63931
:   LENGTH: 957
:   TYPE: PRT
:   ORGANISM: Homo sapiens
: OS-09-724-676-63931

```

18.08; Score 101; DB 5; Length 957;

```

OY 2 QPGPVLSPQACIPING---SSSKQLFSLFHVQPSGNGEKQVT-----TISHSTLT 50
Db 563 QPPPTTSATTALTAIVLSS-----VQRTAGTAATIVTSALQPVLSITLQPIVQG 633
OY 51 IOKGQOKTMPV-----NTIIPTSQFPASILKQITLPGNKIL---SLOASPTQKNRIK 100
Db 634 VGKGQGPPLPYIQQPKPGALIQLNPIQAPVYPVAVPGLPGTALSAVSAQAAAAQKNKIK 693
OY 101 ENVTSCFDE 110
Db 694 EPGGGSFRRD 703

```

RESULT 4

```

Sequence 63931, Application US/09724676A
GENERAL INFORMATION:
APPLICANT: CompuGen LTD
TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 CompuGen
CURRENT APPLICATION NUMBER: US/09/724, 676A
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ. ID NOS: 97222
SOFTWARE: PatentIn version 3.2
SEQ ID NO 63931
LENGTH: 957
TYPE: PRT
ORGANISM: Homo sapiens
US-09-724-676A-63931

```

Query Match	18.0%;	Score 101;	DB 5;	Length 957;
Best Local Similarity	29.2%;	Pred. No. 0.15;		
Matches	38;	Conservative	16;	Mismatches 46;
			Indels	30;
			Gaps	5

```

0Y 2 OPGPLSOPACIPFG-----SSSKOLFSLFHVVOQPSGNGEKVY-----TISHSTLT 50
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 583 QPPPTSOATTAALNVAVLSS-----VQRTAGCTAATVTSALQPPVLSLTQPPVQ 633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
0Y 51 IQCGOKTMY-----NTIIPTSOPPASILKQITLPGNKIL--SLQASPTOKNRK 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 634 VKGQGOPLPYIQQPKRGALIQNLPLQPPYVYVAVPLPGIKALSNVSNQAAAAQKNK 633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
0Y 101 ENVTSCFDE 110
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 694 EPGGGSFRDD 703
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 5
US-09-724-676-63922

```

: Sequence 63922, Application US/09724676
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724,676
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 63922
: LENGTH: 1021
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-724-676-63922

```

Query Match	18.08;	Score 101;	DB 5;	Length 1021;
Best Local Similarity	29.28;	Pred. No. 0.16;		

OY	2	OPGVLSPGPAIPIG	-----SSSKOLFLEHVVOQSGSNGEKT	-----TISHSETL	50
Db	647	QPPPTSOATATLAVVLSS	-----VQPTAGKTAATVTSALQPVLSLTQPTQVG		697
OY	51	IOKGCKRMPT	-----NTIIPTSQFPASILKQILNPGKIL	-----SLQASQNKRIK	100
Db	698	VGKGQGPPIPLYIQGPKRGALIQLNLPQPVYVAVPVLPGTKALSNVSAQAAAAQKNRK			757
OY	101	ENVTSCFDE	110		
Db	758	EPGGGSEFDD	767		

RESULT 6

```

US-09-724-676A-63922
: Sequence 63922, Application US/09724676A
:
: GENERAL INFORMATION:
:
:   APPLICANT: Computegen LTD
:
:   TITLE OF INVENTION: Variants of alternative splicing
:
:   FILE REFERENCE: 129181.4 Computegen
:
:   CURRENT APPLICATION NUMBER: US/09/724, 676A
:
:   CURRENT FILING DATE: 2000-11-28
:
:   NUMBER OF SEQ ID NOS: 97222
:
:   SOFTWARE: PatentIn version 3.2
:
:   SEQ ID NO 63922
:
:   LENGTH: 1021
:
:   TYPE: PRT
:
:   ORGANISM: Homo sapiens
:
US-09-724-676A-63922

```

```

Query Match 18.0% Score 101 DB 5 Length 1021;
Best Local Similarity 29.2% Pred. No. 0.16;
Matches 38; conservative 16; mismatches 46; indels 30; gaps 5

Oy 2 QPQPLVSPACITPG---SSSKQLFSLFRVVOQPSGNGKQVT-----TISHSITLT 50
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 647 QPPPTSATATLATLVVSS-----VORTGKATATVTSALQPVLSLQPTQVG 697

```



```
US-09-724-676-63920
; Sequence 63920, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724, 676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63920
; LENGTH: 711
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-724-676-63920

Query Match
Best Local Similarity 28.9%; Score 93; DB 5; Length 711;
Matches 37; Conservative 15; Mismatches 46; Indels 30; Gaps 5;

QY 2 QPGLVLSQPAIGPTG---SSSKQLFSLFHVVOQPSGNGEKQVT-----TISHSSTLT 50
DB 583 QPPLPSQATLTALTAVLSSS-----VQRTAGKTATVTSALQPPVLSLTQPTQVG 633
QY 51 IOKCGQKTMPIV-----NTIIPTSQFPASILKQITLPGNKIL---SLQASPTOKNRIR 100
DB 634 VGRQGGPTPLVIOQPPKPGALIQLNPLQVPVYKPAVLPGTKALSAVSAQAAAAQKNKKLK 693
QY 101 ENVTSCFR 108
DB 694 EPGGGSFR 701

RESULT 12
US-09-724-676A-63920
; Sequence 63920, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724, 676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63920
; LENGTH: 711
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-724-676A-63920

Query Match
Best Local Similarity 28.9%; Score 93; DB 5; Length 711;
Matches 37; Conservative 15; Mismatches 46; Indels 30; Gaps 5;

QY 2 QPGLVLSQPAIGPTG---SSSKQLFSLFHVVOQPSGNGEKQVT-----TISHSSTLT 50
DB 583 QPPLPSQATLTALTAVLSSS-----VQRTAGKTATVTSALQPPVLSLTQPTQVG 633
QY 51 IOKCGQKTMPIV-----NTIIPTSQFPASILKQITLPGNKIL---SLQASPTOKNRIR 100
DB 634 VGRQGGPTPLVIOQPPKPGALIQLNPLQVPVYKPAVLPGTKALSAVSAQAAAAQKNKKLK 693
QY 101 ENVTSCFR 108
DB 694 EPGGGSFR 701

RESULT 13
US-09-724-676-63923
; Sequence 63923, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
```

```
FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724, 676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63923
; LENGTH: 775
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-724-676-63923

Query Match
Best Local Similarity 28.9%; Score 93; DB 5; Length 775;
Matches 37; Conservative 15; Mismatches 46; Indels 30; Gaps 5;

QY 2 QPGLVLSQPAIGPTG---SSSKQLFSLFHVVOQPSGNGEKQVT-----TISHSSTLT 50
DB 647 QPPLPSQATLTALTAVLSSS-----VQRTAGKTATVTSALQPPVLSLTQPTQVG 697
QY 51 IOKCGQKTMPIV-----NTIIPTSQFPASILKQITLPGNKIL---SLQASPTOKNRIR 100
DB 698 VGRQGGPTPLVIOQPPKPGALIQLNPLQVPVYKPAVLPGTKALSAVSAQAAAAQKNKKLK 757
QY 101 ENVTSCFR 108
DB 758 EPGGGSFR 765

RESULT 14
US-09-724-676A-63923
; Sequence 63923, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724, 676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63923
; LENGTH: 775
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-724-676A-63923

Query Match
Best Local Similarity 28.9%; Score 93; DB 5; Length 775;
Matches 37; Conservative 15; Mismatches 46; Indels 30; Gaps 5;

QY 2 QPGLVLSQPAIGPTG---SSSKQLFSLFHVVOQPSGNGEKQVT-----TISHSSTLT 50
DB 647 QPPLPSQATLTALTAVLSSS-----VQRTAGKTATVTSALQPPVLSLTQPTQVG 697
QY 51 IOKCGQKTMPIV-----NTIIPTSQFPASILKQITLPGNKIL---SLQASPTOKNRIR 100
DB 698 VGRQGGPTPLVIOQPPKPGALIQLNPLQVPVYKPAVLPGTKALSAVSAQAAAAQKNKKLK 757
QY 101 ENVTSCFR 108
DB 758 EPGGGSFR 765

RESULT 15
US-09-724-676-63926
; Sequence 63926, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724, 676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
```

SEQ ID NO 63926
LENGTH: 807
TYPE: PRT
ORGANISM: Homo sapiens
US-09-724-676-63926

Query Match 16.5%; Score 93; DB 5; Length 807;
Best Local Similarity 28.9%; Pred. No. 0.72;
Matches 37; Conservative 15; Mismatches 46; Indels 30; Gaps 5;

OY 2 QPGPVLISQPCIPIC---SSSKOLFSLFHVYQOPSGNEKOVT-----TISHSTLT 50
Db 679 QPPPTSQATTALTAVVLSS-----VQRTAGKTAATVTSALOPPVLSTIQPTQYG 729
OY 51 IQKCGQKTMV-----NTIIPTSQPPASILKQITLPGNKIL---SIQASPTQKNRIK 100
Db 730 VQKQSQPPLVYIQPPKPGALIQLNPLQPVVYKPAVLPGTRALSASVSAQAAAAQKNKIK 789
OY 101 ENVTSCFR 108
Db 790 EPGGGSFR 797

Search completed: February 16, 2003, 22:04:28
Job time : 6.47537 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:55:19 ; Search time 3.72163 Seconds
(without alignments)
2841.441 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_552

Perfect score: 562
Sequence: 1 AQPGLVLSQPAIGPTGSSSK.....ASPTOKNRKIKENTSCFRDE 110

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80.5	14.3	410	2 T50075	probable nicotinat
2	79.5	14.1	3759	2 A35085	trithorax protein
3	78	13.9	346	2 C75572	hypothetical prote
4	76.5	13.6	636	2 S63131	probable membrane
5	76	13.5	403	2 T39846	probable nuclear p
6	74	13.5	434	2 T43197	nucleoporin homolo
7	74	13.2	615	2 JC7576	transcription fact
8	74	13.2	1777	2 T43369	hypothetical prote
9	73	13.0	619	2 A43361	Ets-related transc
10	73	13.0	1466	2 A36426	SPB2 protein - yea
11	72.5	12.9	444	1 C64226	trigger factor Mg2
12	72.5	12.9	545	2 B53309	probable pheromone
13	71.5	12.7	642	2 T39607	fork head protein
14	71.5	12.7	979	2 A35913	regulatory factor
15	70.5	12.5	865	2 AC1966	hypothetical prote
16	70	12.5	218	2 S37658	drought-induced pr
17	70	12.5	441	2 T23461	hypothetical prote
18	70	12.5	688	2 F64111	DNA-directed DNA p
19	70	12.5	1484	2 T42632	breast cancer tumo
20	69.5	12.4	412	1 B70125	hemolysin homolog
21	69.5	12.4	767	2 T19690	hypothetical prote
22	69.5	12.4	1366	2 B86292	F7H2.12 protein -
23	69	12.3	97	2 C22848	hypothetical ORF-2
24	69	12.3	578	2 T48795	origin recognition
25	68.5	12.2	663	2 T40493	hnf-3/forkhead tra
26	68	12.1	379	2 E71296	probable spermidin
27	68	12.1	1858	2 T18273	1 phosphatidylinos
28	67.5	12.0	768	2 T37601	probable transcript
29	67	11.9	366	2 A49076	transcription fact

30	67	11.9	5376	2 T42215	zonadhesin - mouse
31	66.5	11.8	237	1 S73990	ribosomal protein
32	66.5	11.8	343	2 A90654	ATP-binding compon
33	66.5	11.8	343	2 A85505	ATP-binding compon
34	66.5	11.8	510	2 H84824	En/Spn-like transp
35	66.5	11.8	536	2 T17217	hypothetical prote
36	66.5	11.8	1194	2 E96624	hypothetical prote
37	66	11.7	248	2 F96655	hypothetical prote
38	66	11.7	552	2 T24869	probable transcrip
39	66	11.7	753	2 T24869	hypothetical prote
40	66	11.7	1129	2 T19779	hypothetical prote
41	66	11.7	2531	2 A46019	Notch-1 protein -
42	65.5	11.7	247	2 C86204	hypothetical prote
43	65.5	11.7	343	2 G64744	probable ABC-type
44	65.5	11.7	343	2 AC0533	probable ABC trans
45	65.5	11.7	466	2 S45419	hypothetical prote

ALIGNMENTS

RESULT 1
T50075
probable nicotinate phosphoribosyltransferase [imported] - fission yeast (Schizosacch
C:Species: Schizosaccharomyces pombe
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 21-Jul-2000
C:Accession: T50075
R:McDougal, R.C.; Rajandream, M.A.; Barrell, B.G.; Seeger, K.; Harris, D.
submitted to the EMBL Data Library, December 1999
A:Reference number: Z25034
A:Accession: T50075
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-410 <MCD>
A:Cross-references: EMBL:AL133357; PIRN:CA862416.1; GSPDB:GN00066; SPDB:SPAC1486.06
A:Experimental source: Strain 972h(-); cosmid c1486
C:Genetics:
A:Gene: SPDB:SPAC1486.06
A:Map position: 1
C:Superfamily: nicotinate phosphoribosyltransferase

Query Match 14.3%; Score 80.5; DB 2; Length 410;
Best Local Similarity 24.8%; Pred. No. 2.3;
Matches 29; Conservative 21; Mismatches 38; Indels 29; Gaps 5;

QY 15 TGSSSKOLFSEHVQDPSG-----GNEKQVTTISHSTLT----- 51
DB 273 TANSADLANVHGVRQDSCAEYIEKVYKHYSIGVDPSTKVTVHSDALNVRDIELY 332
QY 52 --CRGQRT-MPVNTIIPTSQFPASIIKQITLPGNKILSL-QASPTOKNRKENV 103
DB 333 KYCEKGIKSAFGICITNL-TSDFQVSNPSEVKRMNIVILFSAEGTKAKAISDI 388

RESULT 2
A35085
trithorax protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 20-Sep-1999
C:Accession: A35085; A38240
R:Mazo, A.M.; Huang, D.H.; Mozer, B.A.; David, I.B.
Proc. Natl. Acad. Sci. U.S.A. 87, 2112-2116, 1990
A:Title: The trithorax gene, a trans-acting regulator of the bithorax complex in Dros
A:Reference number: A35085; MUID:90192757; PMID:2107543
A:Accession: A35085
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3759 <MAZ>
R:Mazo, A.M.
submitted to Genbank, January 1990
A:Reference number: A38240
A:Accession: A38240
A:Molecule type: mRNA

A:Residues: 1-2361, 'Y', 2365-2397, 'N', 2399-2405, 'N', 2407-2411, 'N', 2413-3759 <MA2>
A:Cross-references: GB:M31017; NID:g158817; PID:g158818

C:Genetics:

A:Gene: FlyBase:trix
A:Cross-references: FlyBase:FBgn0003862

C:Superfamily: Drosophila trichorax protein

C:Keywords: DNA binding; transcription regulation; zinc finger

Query Match

Best Local Similarity 14.1%; Score 79.5; DB 2; Length 3759;
Matches 30; Conservative 24; Mismatches 50; Indels 61; Gaps 5;

QY 3 PGVLSGPACIPGTSSSKOLFSLFHVVOQPSGNGEKVY-----40

DB 2986 PKVTISQRI-IPAOYQOQOQAOQOMHIPPQOQPLQOQOYVQPSMDITLAEAPVVOQ 3044

QY 41 -----TTTSHSTLTIOKCGQKTPVNTIIPTSQPPASITLKQITLPENKI 86

DB 3045 FVMEPQALBQOELANRQVHSTSSSSSSSSNCISLPTNVNPMQOAPSTTSSSTTRPNNRV 3104

QY 87 LSLQ-----ASPTQKNRIKE-----NVTSCF 107

DB 3105 LPMQORQEPAPLNECPVSSPTPKPEEQPIHQMTASVSKCY 3149

RESULT 3

C75572 hypothetical protein - *Deinococcus radiodurans* (strain R1)

C:Species: *Deinococcus radiodurans*

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000

C:Accession: C75572

R:White, O.; Eilen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

S.; Smith, H.O.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.

A:Reference number: A75250; MIMD:20036986; PMID:10567266

A:Accession: C75572

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-346 <WHI>

A:Cross-references: GB:AE001864; GB:AE000513; NID:g6457659; PIDN:AAF09603.1; PID:g645766

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0008

A:Map position: 1

Query Match 13.9%; Score 78; DB 2; Length 346;
Best Local Similarity 22.0%; Pred. No. 3.3;
Matches 24; Conservative 24; Mismatches 45; Indels 16; Gaps 2;

QY 9 QPAGIPGTSSSKOLFSLFHVVOQPSGNGEKVYTTTSHSTLTIOKCGQK-----HSTLTIO 52

DB 175 RPSEVQATAPSRVILNQVTRVTSVPDLDPGEERAPLIALDSGRVVEPTLHPASVQ 234

QY 53 KCGQKTPVNTIIPTSQPPASITLKQITLPENKIISQASPTQKNRIKE 101

DB 235 RLDITLPAKTLPVVLGAPPASLRVOSQTLQPTRVVAAPELLGLRL 283

RESULT 4

S63131 probable membrane protein YNL176c - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein N161

C:Species: *Saccharomyces cerevisiae*

C:Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 19-Apr-2002

C:Accession: S63131

R:Obermaier, B.; Piravandi, E.; Rinke, M.; Dondoy, H.

submitted to the Protein Sequence Database, April 1996

A:Reference number: S63132

A:Accession: S63131

A:Molecule type: DNA

A:Residues: 1-636 <OBE>

A:Cross-references: EMBL:Z11452; NID:g1302153; PID:e239549; PID:g1302154; GSPDB:GN000

A:Experimental source: strain S286C

C:Genetics:

A:Gene: MIPS:YNL176c

A:Cross-references: SGD:S0005120

A:Map position: 14L

C:Keywords: transmembrane protein

F:464-480/Domain: transmembrane #status predicted <TM>

Query Match

Best Local Similarity 13.6%; Score 76.5; DB 2; Length 636;
Matches 23; Conservative 12; Mismatches 34; Indels 7; Gaps 2;

QY 7 LSPAGIPGTSSSKOLFSLFHVVOQPSGNGEKVYTTTSHSTLTIOKCGQKTPVNTIIP 66

DB 296 ISVPTSSSVSSSSSKVPS-----NRPSSSSSDDTTSAVSTYFGLQSTTS--SSIPP 348

QY 67 TSQPPASITLKQITLP 82

DB 349 TTOPSTSTISTSPDP 364

RESULT 5

T39846 probable nuclear pore protein - fission yeast (*Schizosaccharomyces pombe*)

C:Species: *Schizosaccharomyces pombe*

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T39846

R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Devlin, K.; Churcher, C.M.

submitted to the EMBL Data Library, February 1998

A:Reference number: Z21884

A:Accession: T39846

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-403 <LYN>

A:Cross-references: EMBL:AL021839; PIDN:CAA17059.1; GSPDB:GN00067; SPDB:SPBC1967.15

A:Experimental source: strain 972n; cosmid c1967

C:Genetics:

A:Gene: SPDB:SPBC1967.15

A:Map position: 2

A:Introns: 6/7

Query Match 13.5%; Score 76; DB 2; Length 403;
Best Local Similarity 27.0%; Pred. No. 6.3;
Matches 30; Conservative 14; Mismatches 49; Indels 18; Gaps 3;

QY 11 AGIPGTSSSKOLF--SLFHVVOQPSGNGEKVYTTTSHSTLTIOKCGQK-----58

DB 50 AGPPTGSSSAPPFGNSIFGRTQOQPTTSFSNTTTNAGQSTVFGONASRTGNSNTOPLFS 109

QY 59 -----MPVNTIIPTSQPPASITLKQITLPENKIISQASPTQKNRIKENY 103

DB 110 WSTVNNPTKRVDETNATIPSSLSSGISPNATVSNQYGPAPPSVEEY 160

RESULT 6

T43197 nucleoporin homolog - fission yeast (*Schizosaccharomyces pombe*) (fragment)

C:Species: *Schizosaccharomyces pombe*

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000

C:Accession: T43197

R:Toshioaka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.

DNA Res. 4, 363-369, 1997

A:Title: Identification of open reading frames in *Schizosaccharomyces pombe* cDNAs.

A:Reference number: Z17323; MIMD:98162722; PMID:9501991

A:Accession: T43197

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-434 <YOS>

A:Cross-references: EMBL:DB9269; NID:g1749745; PIDN:BAAL3930.1; PID:g1749746

A:Experimental source: strain PR745

Query Match

13.5%; Score 76; DB 2; Length 434;

Best Local Similarity 27.0%; Pred. No. 6.9;
Matches 30; Conservative 14; Mismatches 49; Indels 18; Gaps 3;
Db 11 AGITPGSSSKOLF---SLFHYVQPSGSGNEKOVTTISHSTLTIOKCGOKT----- 58
11 AGITPGSSSKOLF---SLFHYVQPSGSGNEKOVTTISHSTLTIOKCGOKT----- 58
Db 81 AGITPGSSSKOLF---SLFHYVQPSGSGNEKOVTTISHSTLTIOKCGOKT----- 58
59 -----MPVNTIIPTSOPFPASILKQITL-PGNKILSLQASPTOKNRKENY 103
141 MSTVNNPTKPPDEMTNATIPSSILLSSGISPMATSNMAYGAPPSVEQY 191

RESULT 7
JC7576
transcription factor Elf-1, type 1 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 30-Jun-2001
C:Accession: J07576
R:Nishiyama, C.; Takahashi, K.; Nishiyama, M.; Okumura, K.; Ra, C.; Ohtake, Y.; Yokota, Y.
Biochem. Biotechnol. Biochem. 64, 2601-2607, 2000
A:Title: Splice isoforms of transcription factor Elf-1 affecting its regulatory function
A:Reference number: J07576; MUID: 21077473; PMID:11210123
A:Contents: Mast cell line, RBL-2H3
A:Accession: J07576
A:Molecule type: mRNA
A:Residues: 1-615 <NTS>
A:Cross-references: DDBJ:AB030215
C:Comment: This protein, as a key transcription factor for immune-related genes, has the
C:Genetics:
A:Gene: Elf-1

Query Match 13.2%; Score 74; DB 2; Length 615;
Best Local Similarity 30.9%; Pred. No. 17;
Matches 30; Conservative 11; Mismatches 38; Indels 18; Gaps 5;
Db 2 QPCGVLS--QPAGIPTGSSSKOLFSLFHYVQPSGSGNEKOVTTISHSTLTIOKCGOKT 59
362 QPSEVLRFTVQPSQAPYPT---QLFRVHVVPQVQAVPEARIT---STMDEAAGSSVQ 414
QY 60 PVNTIIPTSOPFPASILKQITLPGNKIL---SLQASP 93
Db 415 GRTIQAFTQVPPV---VVVSPGNQQLHTVTLQTPV 445

RESULT 8
T34369
hypothetical protein T19D12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34369
R:Favella, A.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid T19D12.
A:Reference number: Z21513
A:Accession: T34369
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1777 <PAV>
A:Cross-references: EMBL:U41263; PIDN:AAC24428.1; GSPDB:GN00020; CESP:T19D12.1
C:Experimental source: strain Bristol N2; clone T19D12
C:Genetics:
A:Gene: CESP:T19D12.1
A:Map position: 2
A:Introns: 36/1; 134/2; 180/1; 622/3; 691/2; 754/1; 1111/2; 1174/1; 1271/3; 1322/2; 1681

Query Match 13.2%; Score 74; DB 2; Length 1777;
Best Local Similarity 27.2%; Pred. No. 58;
Matches 34; Conservative 18; Mismatches 41; Indels 32; Gaps 6;
QY 1 AOPPVLSOPGIPFGSSSKOLFSLFHYVQPSGSGNEKOVTTISHSTLTIOKCG----- 55
Db 928 AOPASSTSTMAITVTVSS-----QSPSTSPAQSSSTPACSSITVTVQSSSFOSP 977

QY 56 QKMPV--NTIIPTSOPFPASI-----LKOITLPG--NKILSLQASPTOKNRK 100
Db 978 OSTQIGSSITVPSQAPSSSTSGPPTTQICPNQOTFKGQGVYKEMLPSTQNMALNA 1037
QY 101 --ENV 103
Db 1038 FVENY 1042

RESULT 9
A43361
Ets-related transcription factor Elf-1 - human
N:Alternate names: E74-like factor Elf-1
C:Species: Homo sapiens (man)
C:Date: 11-Feb-1993 #sequence_revision 11-Feb-1993 #text_change 07-May-1999
C:Accession: A43361; A42122
R:Leiden, J.M.; Wang, C.Y.; Petryniak, B.; Markovitz, D.M.; Nabel, G.J.; Thompson, C.
J. Virol. 66, 5890-5897, 1992
A:Title: A novel Ets-related transcription factor, Elf-1, binds to human immunodeficiency
A:Reference number: A43361; MUID:92407982; PMID:1527846
A:Accession: A43361
A:Molecule type: mRNA
A:Residues: 1-619 <LEI>
A:Cross-references: GB:M82882
R:Thompson, C.B.; Wang, C.Y.; Ho, I.C.; Bohjanen, P.R.; Petryniak, B.; June, C.H.; M.
Mol. Cell. Biol. 12, 1043-1053, 1992
A:Title: cis-acting sequences required for inducible interleukin-2 enhancer function
A:Reference number: A42122; MUID:92186836; PMID:1545787
A:Accession: A42122
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 204-282, 'G', 284-289 <THO>
A:Experimental source: T-cells
A:Note: sequence extracted from NCBI backbone (NCBI:88288)
C:Genetics:
A:Gene: GDB:ELF1
A:Cross-references: GDB:131648
A:Map position: 1p36-1p36
C:Superfamily: ets DNA-binding domain homology
C:Keywords: DNA binding; transcription regulation
F:210-290/Domain: ets DNA-binding domain homology <ETS>

Query Match 13.0%; Score 73; DB 2; Length 619;
Best Local Similarity 28.9%; Pred. No. 21;
Matches 37; Conservative 11; Mismatches 50; Indels 30; Gaps 6;

QY 1 AOPGVLS--QPAGIPTGSSSKOLFSLFHYVQ---PSGNEKOVTTISHSTLTIOK 54
Db 362 AOPSEVLRFTVQPSQAPYPT---QLFRVHVVPQVQAVPEG--EAARTSTMODETLNSSVQ 416
QY 55 GOKTNPVNTIIPTSOPFPASILKQITL-----PGNKILSLQASPTOK 96
Db 417 SIRTQAFTQVPPV---VVVSPGNQQLHTVTLQTPVLTIVIASTDPSAGTSOKFTLQALPSSQ 476

QY 97 -NRKENY 103
Db 477 MTVLKENY 484

RESULT 10
A36426
SPA2 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein L1209; protein YIL021w
C:Species: Saccharomyces cerevisiae
C:Date: 12-Apr-1991 #sequence_revision 12-Apr-1991 #text_change 21-Jul-2000
C:Accession: A36426; S64769; S69386
R:Gehrung, S.; Snyder, M.
J. Cell Biol. 111, 1451-1464, 1990
A:Title: The SPA2 gene of Saccharomyces cerevisiae is important for pheromone-induced
A:Reference number: A36426; MUID:91009481; PMID:2211820
A:Accession: A36426
A:Molecule type: DNA
A:Residues: 1-1466 <GEH>

A:Cross-references: GB:X53731; NID:g4519; PIDN:CAA37763.1; PID:g4520
R:Goffeau, A.; Purnelle, B.
Submitted to the Protein Sequence Database, May 1996
A:Reference number: S64761
A:Accession: S64761
A:Molecule type: DNA
A:Residues: 1-1466 <GOR>
A:Cross-references: EMBL:Z73126; NID:g1360195; PID:e245456; PID:g1360196; MIPS:YLL021w
A:Note: experimental_source strain S288C
R:Purnelle, B.; Goffeau, A.
Submitted to the EMBL Data Library, April 1996
A:Description: The sequence of 32 kb on the left arm of yeast chromosome XII reveals 14
mily and a new ABC transporter homologous to the human multidrug resistance protein.
A:Reference number: S69380
A:Accession: S69386
A:Molecule type: DNA
A:Residues: 1-1466 <PUS>
A:Cross-references: EMBL:X97560; NID:g1297003; PID:e238713; PID:g1297010
C:Genetics:
A:Gene: SGD:SPA2
A:Cross-references: SGD:S0003944; MIPS:YLL021w
A:Map position: 12L
C:Keywords: coiled coil

Query Match 13.0%; Score 73; DB 2; Length 1466;
Best Local Similarity 23.1%; Pred. No. 58;
Matches 28; Conservative 27; Mismatches 32; Indels 34; Gaps 6;

OY 5 PVL-SQAPGIPGSS--SSKQFLSLF--HVVOOPSGGNE-----KQVTTISHST 48
DB 600 PIRPSSNGVPTTNRKPSGTGLSLMIDSSIKSHKDNKYSPILAVTSASASAS 659

OY 49 LTIQKCGQKTMF-----VNTIIPISQPPPSILKQITLPGKILSLQASPTQKNRIKENT 104
DB 660 SNISEIPLKTLPPQAKIGVIPPSE-----NQVNPINIENTEDNKRSDIT 705

OY 105 S 105
DB 706 N 706

RESULT 11

C64226
triglycer factor MG238 (similarity) - Mycoplasma genitalium
N:Contents: peptidylprolyl isomerase (EC 5.2.1.8), ribosome-bound
C:Species: Mycoplasma genitalium
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Sep-2000
C:Accession: C64226
R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;
M.; Fuhmann, J.; Nguyen, D.; Usterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.
C.A.; Venter, J.C.
Science 270, 397-403, 1995
A>Title: The minimal gene complement of Mycoplasma genitalium.
A:Reference number: A64200; MUID:96026346; PMID:7569993
A:Accession: C64226
A:Molecule type: DNA
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Genetics:
A:Gene: SGC3
A:Cross-references: GB:U039701; GB:I43967; NID:g1045915; PID:g1045927; TIGR:MG238
A:Experimental source: strain G-37

Query Match 12.9%; Score 72.5; DB 1; Length 444;
Best Local Similarity 26.1%; Pred. No. 16;
Matches 31; Conservative 18; Mismatches 31; Indels 39; Gaps 6;

OY 10 PACIPTSSSKQLFLFHVVOOPSGGNEKQVTTISHSTLTQKCGQKTMPTVNTIPTSQ 69
DB 267 PPOFGFTSLKAFKSYFHILME-----NKKQETILQENN-----QKINQPLT-TTKRP--- 314

OY 70 FPPASITKQITLPGKILSLQASPTQKNRI-----KENVT 104
DB 315 FLPEALIK---LEARNILKLQSQAEQYKIPFEKLASNTITLQDRNIKEKENV 370

RESULT 12

B53309
Probable pheromone binding protein prg2 - Enterococcus faecalis plasmid pCF10
N:Alternate names: pheromone responsive gene 2 protein
C:Species: Enterococcus faecalis
C>Date: 13-Sep-1995 #sequence_revision 13-Sep-1995 #text_change 20-Aug-1999
C:Accession: B53309
R:Ruhfel, R.E.; Mantas, D.A.; Dunny, G.M.
J. Bacteriol. 175, 5253-5259, 1993
A>Title: Cloning and characterization of a region of the Enterococcus faecalis conj
A:Reference number: A53309; MUID:93352432; PMID:8349565
A:Accession: B53309
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-545 <RUH>
A:Cross-references: GB:L14285; NID:g309660; PIDN:AAA2554.1; PID:g309662
C:Genetics:
A:Gene: prg2
A:Genome: plasmid

C:Superfamily: dipeptide transport protein

Query Match 12.9%; Score 72.5; DB 2; Length 545;
Best Local Similarity 26.1%; Pred. No. 20;
Matches 23; Conservative 15; Mismatches 29; Indels 21; Gaps 3;

OY 15 TGSSSKQFLFHVVOOPSGGNEKQVTTISHSTLTQKCGQKTMPTVNTIPTSQPPAS 74
DB 135 TVSPNVELFSAIRNAKIASGQAK-----DLAVKSIQETLTELVEPPYF---- 183

OY 75 ILKQITLPGKILSLQASPTQKNRIE 101
DB 184 -----TDLSTLTATYTPVQQAIKE 202

RESULT 13

T39607
fork head protein type transcription factor - fission yeast (Schizosaccharomyces pombe
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T39607
R:Lyne, M.; Wood, V.; Rajandream, M.A.; Barrett, B.G.; Beck, A.; Reinhardt, R.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z21866
A:Accession: T39607
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-642 <LYN>
A:Cross-references: EMBL:AL023554; PIDN:CAA19034.1; GSPDB:GN00067; SPDB:SPBC1665.15C
A:Experimental source: strain 972h-; cosmid c1665
C:Genetics:
A:Gene: SPDB:SPBC1665.15C
A:Map position: 2
A:introns: 171/2

Query Match 12.7%; Score 71.5; DB 2; Length 642;
Best Local Similarity 28.3%; Pred. No. 31;
Matches 28; Conservative 15; Mismatches 39; Indels 17; Gaps 4;

OY 17 SSSKQFLFHVVOOP-----SGGNEKQV-----TTISHSTLTQKCGQKTMPTVNT 63
DB 480 SSPKTIQDEQLHLDQGVLDVNGFEGISSWRESVYVNTKLSSVDSPTVNLANSNSKSSPV-A 538

OY 64 IIPISQPPASITKQITLPGKILSLQASPTQKNRIKEN 102
DB 539 VQVSTLTPQASANKQAKEMESK---MSNSPTQSKTEEN 574

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:54:50 ; Search time 2.11991 Seconds

(Without alignments)
2152.161 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_552

Perfect score: 562
Sequence: 1 AOPGPVLSQAPAGIPTGSSSK.....ASPTQKRIKENYTCSEFDE 110

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	562	100.0	801	T2DT_HUMAN	O92750 homo sapien
2	92.5	16.5	1083	T2D3_HUMAN	O00268 homo sapien
3	80.5	14.3	410	NPTI_SCHPO	Q06133 schizosacch
4	79.5	14.1	3726	TRX_DROME	P20659 drosophila
5	76.5	13.6	612	ELFI_MOUSE	O60775 mus musculu
6	76.5	13.6	636	YNR6_YEAST	P53882 saccharomyc
7	73	13.0	619	ELFI_HUMAN	P32519 homo sapien
8	73	13.0	1466	SPA2_YEAST	P23201 saccharomyc
9	72.5	12.9	444	TIG_MYCE	P47480 mycoplasma
10	71.5	12.7	979	REF1_HUMAN	P22670 homo sapien
11	71	12.6	528	PODX_HUMAN	O00592 homo sapien
12	71	12.5	1531	NFT5_HUMAN	O94916 homo sapien
13	70	12.5	688	DP3X_HAEN	P43746 haemophilus
14	69	12.3	551	SMA4_MOUSE	P97471 mus musculu
15	68.5	12.2	539	RGSK_MOUSE	O99q21 mus musculu
16	68.5	12.2	531	RN27_MOUSE	O99q21 mus musculu
17	68.5	12.2	663	SEPI_SCHPO	O43058 schizosacch
18	68.5	12.2	2300	CYAA_NEUCR	O01631 neurospora
19	68	12.1	552	SMA4_RAT	O70437 rattus norv
20	68	12.1	1858	PKA2_DICDI	P54674 dictyostell
21	67.5	12.0	2442	CBP_HUMAN	O92763 homo sapien
22	67	11.9	366	T2AA_DROME	P52654 drosophila
23	67	11.9	552	SMA4_PIG	O9qk69 sus scrofa
24	67	11.9	5376	ZAN_MOUSE	O88739 mus musculu
25	66.5	11.8	237	RL23_MYCPN	P75578 mycoplasma
26	66.5	11.8	239	CU58_HUMAN	P58505 homo sapien
27	66	11.7	194	RS7_METVA	P14037 methanococ
28	66	11.7	554	SMA4_HUMAN	O13485 homo sapien
29	66	11.7	2531	NTC1_MOUSE	O01705 mus musculu
30	65.5	11.7	343	ABC_ECOLI	P30750 escherichia
31	65.5	11.7	466	YB16_YEAST	P38177 saccharomyc
32	65.5	11.7	628	ABFA_ASFNG	P42234 aspergillus
33	65.5	11.7	983	Y144_HUMAN	Q14157 homo sapien

34	65	11.6	284	1	SIC1_YEAST	P38634 saccharomyc
35	65	11.6	448	1	YAP1_CHICK	P46936 gallus galli
36	65	11.6	545	1	AGAL_ASPNG	P28351 aspergillus
37	65	11.6	896	1	EP15_HUMAN	P42566 homo sapien
38	65	11.6	1228	1	ECM_HUMAN	O13201 homo sapien
39	65	11.6	1273	1	WEB1_YEAST	P38968 saccharomyc
40	64.5	11.5	2441	1	CBP_MOUSE	P45481 mus musculu
41	64	11.4	199	1	SOXE_SULAC	O53765 sulfolobus
42	64	11.4	234	1	TNEA_CAPII	P13296 capra hircu
43	64	11.4	322	1	YD67_MYCPN	P75414 mycoplasma
44	64	11.4	326	1	VS09_ROTBT	P30210 bovine tota
45	64	11.4	677	1	VGP_EBORS	O89853 ebola virus

ALIGNMENTS

```

RESULT 1
T2DT_HUMAN          STANDARD:      PRT;      801 AA.
ID                  092750:
AC 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE Transcription Initiation factor TFIID 105 kDa subunit (TAFII-105)
DE (TAFII105) (Fragment).
GN TAF4B OR TAF2C2 OR TAFII105.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RX MEDLINE=97011146; PubMed=8858156;
RA Dikstein R., Zhou S., Tjian R.;
RT "Human TAFII 105 is a cell type-specific TFIID subunit related to
RL hTAFII130."
CC -1- FUNCTION: CELL TYPE-SPECIFIC SUBUNIT OF TFIID THAT MAY FUNCTION AS
CC A GENE-SELECTIVE COACTIVATOR IN CERTAIN CELLS. TFIID IS A
CC MULTIMERIC PROTEIN COMPLEX THAT PLAYS A CENTRAL ROLE IN MEDIATING
CC PROMOTER RESPONSES TO VARIOUS ACTIVATORS AND REPRESSORS.
CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs). TAFII105 MAY FORM
CC HETERODIMERS WITH TAFII130.
CC -1- SUBCELLULAR LOCATION: Nucleus.
CC -1- TISSUE SPECIFICITY: FOUND IN B CELLS BUT NOT IN OTHER TESTED
CC CELLS.
CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; Y09321; CA70499.1; -
CC Genew; HGNC:11538; TAF4B.
CC MIM; 601689; -
CC InterPro; IPR003894; TAF_hom.
CC SMART; SM00549; TAFH; 1.
CC KW Transcription regulation; Nuclear protein.
CC FT NON_TER
CC SQ SEQUENCE 801 AA; 85658 MW; D12B4933FEA49CD2 CRC64;
Query Match 100.0%; Score 562; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 3 5e-48;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 AOPGPVLSQAPAGIPTGSSSKQLFSLFHVVOOPSGGENKQVTTISHSTLTIOKCGOKTTP 60
|||||

```

DB 443 ACPGVLSPAGIPGSSSKQLFSLFHVQOPSGNGKQVTTISHSSTLTITQKCGOKTWP 502

QY 61 VNTIIPSPASILKQITLPGKILSLASPTQKRIKENTVSCFRDE 110

DB 503 VNTIIPSPASILKQITLPGKILSLASPTQKRIKENTVSCFRDE 552

RESULT 2

T2D3_HUMAN

ID T2D3_HUMAN STANDARD; PRT; 1083 AA.

AC 000268; 099721; 09BX42; 09BR40;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUN-1998 (Rel. 36, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Transcription initiation factor TFIID 135 kDa subunit (TAFII-135)

DE (TAFII135) (TAFII-130) (TAFII130).

GN TAF4 OR TAF4A OR TAF2C1 OR TAFII135 OR TAFII130.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OC NCBI_TaxID=9606;

OX [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-97336072; PubMed-9192867;

RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R., Jones M., Scavrides G., Almeida J.P., Babbage A.K., Bagguley C.L., Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M., Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J., Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M., Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R., Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M., Ellington A.G., Frankland J.A., Fraser A., French L., Garner P., Hammond D.V., Griffiths C., Griffiths M.N.D., Gilliam J.R., Hall R.E., Hameed S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J., Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D., Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Taylor S., Leharstain H.H., Leverisa M.A., Lloyd C., Lloyd D.M., Lovell J.D., Marsh S.A., Martin S.L., Mooney M.J.F., Mullikin J.C., Nickerson T., Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I., Phillimore B.J.C.T., Prithalingam S.R., Plumb R.W., Ramsay H., Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S., Skuce C.D., Smith M.L., Soderlund C., Stewart C.A., Sulston J.E., Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A., Tracey A., Troman A.C., Vaudin M., Wall M., Wallis J.M., Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A., Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S., Rogers J.

RA The DNA sequence and comparative analysis of human chromosome 20.*

RT Nature 414:863-871 (2001).

RL [3]

RN SEQUENCE OF 105-1083 FROM N.A. AND PARTIAL SEQUENCE.

RP MEDLINE-97098442; PubMed-8942982;

RX Tanese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.

RT Molecular cloning and analysis of two subunits of the human TFIID complex: hTAFII130 and hTAFII100.*

RT Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616 (1996).

CC -1- FUNCTION: MAKES PART OF TFIID IS A MULTISUBUNIT PROTEIN COMPLEX THAT PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION BY THE AF-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.

CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: Y11354; CAA72189.1; -

DR EMBL: AL137077; CAC36006.1; -

DR EMBL: AL109911; CAC22312.2; -

DR EMBL: U75308; AAC50901.1; -

DR TRANSFAC: T02328; -

DR Genew; HGNC: 11537; TAF4.

DR MIM: 601796; -

DR InterPro: IPR003894; TAF_hom.

DR SMART: SM00549; TAFH.1.

KW Transcription regulation; Nuclear protein.

FT DOMAIN 39 42 POLY-HIS.

FT DOMAIN 52 57 POLY-ALA.

FT DOMAIN 98 101 POLY-GLY.

FT DOMAIN 142 148 POLY-ALA.

FT DOMAIN 268 275 POLY-PRO.

FT DOMAIN 331 337 POLY-ALA.

FT DOMAIN 680 683 POLY-PRO.

FT DOMAIN 808 813 POLY-ALA.

FT DOMAIN 828 831 POLY-ASP.

FT CONFLICT 105 117 PCPSPRRRLVPA -> GGLGLQQRGREGS (IN REF. 3).

FT CONFLICT 136 136 A -> S (IN REF. 2).

FT CONFLICT 185 185 G -> GPG (IN REF. 2).

FT CONFLICT 233 264 MISSING (IN REF. 3).

FT CONFLICT 293 293 P -> L (IN REF. 3).

SO SEQUENCE 1083 AA; 109943 MW; A645827572A0752 CMC64;

Query Match 16.58; Score 92.5; DB 1; Length 1083;

Best Local Similarity 27.38; Pred. No. 0.2;

Matches 33; Conservative 19; Mismatches 50; Indels 19; Gaps 5;

QY 2 QPGVLS--OPAGIPGSSSKQLFSLFHVQOPSGNGKQVTTISHSSTLTITQKCGOKTWP 59

DB 716 QP-VLSLTPGTVGVGKQOPRRLVQOPRPFALIRPOVTLTPYVALNQ----- 768

QY 60 PVNTIIPSPASILKQITLPGKILSLASPTQKRIKENTVSCFRD 109

DB 769 PNRIMLTTPQQLQNLNLPVVPVPRVLPRTALSAVSAQAAAKNKIKERGGSGFRD 828

QY 110 E 110

DB 829 D 829

RESULT 3

NP1L_SCHPO

ID NP1L_SCHPO STANDARD; PRT; 410 AA.

AC 090TK3;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Probable nicotinate phosphoribosyltransferase (EC 2.4.2.11)

DE (NAPRTase).

GN SPAC1486.06.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomycetes.

OC NCBI_TaxID=4896;

OX [1]

RP SEQUENCE FROM N.A.

RX STRAIN-972;

RX MEDLINE-21848401; PubMed-11859360;

RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA Jones K., Jones L., Jones M., Leather S., McDonald S., Melan J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Mblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skellon J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tavey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Wolckaert G., Aert R., Robben J., Grymoprez B.,
RA Welljens I., Vanstreels E., Rieger M., Schefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Filtz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lebrach R., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Wambutt R., Punnett B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leclaire V., Mottler S.,
RA Gallbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Rhode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerretti L., Love T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -1- CATALYTIC ACTIVITY: Nicotinate D-ribonucleotide + diphosphate -
CC nicotinate + 5-phospho-alpha-D-ribose 1-diphosphate.
CC -1- SIMILARITY: BELONGS TO THE NMRASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AL133357; CAB62416.1; -
KW Pyridine nucleotide biosynthesis; Transferase; Glycosyltransferase.
SQ SEQUENCE 410 AA; 46694 MW; EA230MDPD9ABA6077 CAC64;
OY Query Match 14.3%; Score 80.5; DB 1; Length 410;
Best Local Similarity 24.8%; Pred. No. 0.98;
Matches 29; Conservative 21; Mismatches 38; Indels 29; Gaps 5;
OY 15 TGSSSKOLFSLFHVYQPSG-----GNEKVTYISHSTLT----- 51
DB 273 TANSADDLANVFHGVRQDSQCAEYIEKVKHYSIGVDPSTKVTYHSDALNVDCRIELY 332
OY 52 --QKSGQKT-MPVNTIITPSPPASILIKOTLTPGNKILSL-QASPTQKNRIKENV 103
DB 333 KYCEKCGIKSAFGICITNL-TSDPKVNSPSEVKPMNIYIKLFSAGTKRAVTSDDI 388
RESULT 4
TRX_DROME STANDARD: PRT: 3726 AA.
ID AC P20659; Q27255; Q27327;
DT 01-FEB-1991 (Rel. 17, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trithorax protein.
GN TRX.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=90192757; PubMed=2107543;
RX

RA Mazo A.M., Huang D.-H., Mozer B.A., David I.B.;
RT "The trithorax gene, a trans-acting regulator of the bithorax complex
RT in Drosophila, encodes a protein with zinc-binding domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2112-2116(1990).
[2]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND CHARACTERIZATION.
RX MEDLINE=95009521; PubMed=7924996;
RA Sedkov Y., Tiliib S., Mizrokh L., Mazo A.;
RT "The bithorax complex is regulated by trithorax earlier during
RT Drosophila embryogenesis than is the Antennapedia complex, correlating
RT with a bithorax-like expression pattern of distinct early trithorax
RT transcripts.";
RL Development 120:1907-1917(1994).
[3]
RP SEQUENCE FROM N.A.
RX STRAIN=Oregon-R;
RA MEDLINE=96100387; PubMed=8555104;
RT Tiliib S., Sedkov Y., Mizrokh L., Mazo A.;
RT "Conservation of structure and expression of the trithorax gene
RT between Drosophila virilis and Drosophila melanogaster.";
RL Mech. Dev. 53:113-122(1995).
[4]
RP CHARACTERIZATION.
RX MEDLINE=95047388; PubMed=7958911;
RA Kuzin B., Tiliib S., Sedkov Y., Mizrokh L., Mazo A.;
RT "The Drosophila trithorax gene encodes a chromosomal protein and
RT directly regulates the region-specific homeotic fork head.";
RL Genes Dev. 8:2478-2490(1994).
CC -1- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION
CC WITH GENES OF BITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.
CC IT CAN BEHAVE AS AN ACTIVATOR OF BX-C.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- MISCELLANEOUS: THIS PROTEIN HAS BEEN EXPERIMENTALLY SHOWN TO BIND
CC ZINC.
CC -1- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC -1- SIMILARITY: CONTAINS 5 PHD-TYPE ZINC FINGERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M31617; AAA29025.1; -
DR EMBL: Z50152; CAA90514.1; -
DR EMBL: Z50152; CAA90513.1; -
DR EMBL: Z31725; CAA83516.1; -
DR EMBL: Z31725; CAA83515.1; -
DR PIR: A35085; A35085.
DR HSSP: P20393; 1A6Y.
DR TRANSFAC: T00850; -
DR FLYBase: FBgn003862; trx.
DR InterPro: IPR003889; FYRIC_N.
DR InterPro: IPR003888; FYRIC_C.
DR InterPro: IPR003616; PostSet.
DR InterPro: IPR001214; SET.
DR InterPro: IPR001965; znf_PHD.
DR InterPro: IPR001841; znf_ring.
DR Pfam: PF00628; PHD; 3.
DR Pfam: PF00856; SET; 1.
DR SMART: SM00542; FYRC; 1.
DR SMART: SM00541; FYRN; 1.
DR SMART: SM00249; PHD; 4.
DR SMART: SM00508; PostSet; 1.
DR SMART: SM00164; RING; 3.
DR SMART: SM00317; SET; 1.
DR PROSITE: PS0280; SET; 1.
DR PROSITE: PS01359; ZF_PHD_1; 4.

DR PROSITE: PS50016; 2F, PHD_2; 3.
 KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
 Nuclear protein; Developmental protein; Activator;
 Alternative splicing.
 FT ZN_FING 1266 1347 PHD-TYPE 1.
 FT ZN_FING 1348 1393 PHD-TYPE 2.
 FT ZN_FING 1421 1482 PHD-TYPE 3.
 FT ZN_FING 1734 1793 PHD-TYPE 4 (ATYPICAL).
 FT ZN_FING 1794 1844 PHD-TYPE 5 (ATYPICAL).
 FT DOMAIN 3599 3708 SET.
 FT DOMAIN 512 516 POLY-SER.
 FT DOMAIN 565 570 POLY-ASP.
 FT DOMAIN 661 664 POLY-SER.
 FT DOMAIN 905 910 POLY-SER.
 FT DOMAIN 1576 1582 POLY-GLN.
 FT DOMAIN 2298 3027 GLN-RICH.
 FT DOMAIN 3032 3040 POLY-SER.
 FT DOMAIN 3181 3184 POLY-GLN.
 FT DOMAIN 3220 3225 POLY-GLN.
 FT VARSPLIC 1 368 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 2023 2025 P -> PWLSPFLKFLGSLTHGGLLWLLGCVVRLKQCG
 (IN REF. 1).
 FT CONFLICT 2341 2341 R -> S (IN REF. 1).
 FT CONFLICT 2392 2392 G -> S (IN REF. 1).
 SQ SEQUENCE 3726 AA; 400575 MM; D275650763D1CF5 CRC64;

Query Match 14.18; Score 79.5; DB 1; Length 3726;
 Best Local Similarity 18.2%; Pred. No. 17;
 Matches 30; Conservative 24; Mismatches 50; Indels 61; Gaps 5;

QY 3 PGPLSPAGIPGTSSSKOL-FSLFHVQOPSGNEKOV----- 40
 DB 2953 PKYTISQQR-IPACTQOOQLQAOAMHIFPOQOPPLQOOOVOPSPMTITLAEPPVQSO 3011
 QY 41 -----TTTISHSTLTITQKCGKTMPEVNTITPTSPASIIKQITLPQNKI 86
 DB 3012 FVMEPALQEOELANRYQHFSTSSSSSSNCLEPTVNVNMQQAPSTTSSSTRPNNRV 3071
 QY 87 LSLQ-----ASPTQKNRIKE-----NVTSCF 107
 DB 3072 LPMQORQEPAPLNECPVSSPTPPKPEQPIIHQMTSASVSKCY 3116

RESULT 5

ELF1_MOUSE
 ID ELF1_MOUSE STANDARD; PRT; 612 AA.
 AC Q60775;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE ETS-related transcription factor Elf-1 (E74-like factor 1).
 GN ELF1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fibroblast;
 RA MEDLINE=96257231; PubMed=8666284;
 RA Davis J.N., Roussee M.F.;
 RT "Cloning and expression of the murine Elf-1 cDNA."
 RL Gene 171:265-269(1996).
 CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT APPEARS TO BE REQUIRED FOR THE
 T-CELL-RECEPTOR-MEDIATED TRANS ACTIVATION OF HIV-2 GENE
 EXPRESSION. BINDS SPECIFICALLY TO TWO PURINE-RICH MOTIFS IN THE
 HIV-2 ENHANCER. ELF-1 BINDS TO THE UNDERPHOSPHORYLATED FORM OF RB.
 CC MAY INTERACT WITH OTHER TRANSCRIPTION FACTORS IN ORDER TO REGULATE
 SPECIFIC GENES (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY FOUND IN HEMATOPOIETIC CELLS.
 CC DETECTED IN OTHER CELL TYPES SUCH AS FIBROBLASTS.
 CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: U19617; AAB17097.1; -.
 DR HSBP: P28324; IBC8.
 DR TRANSFAC: T05012; -.
 DR MGD: MGI:107180; Elf1.
 DR InterPro: IPR000418; Ets.
 DR InterPro: IPR002341; HSF_ETs.
 DR Pfam: PF00178; Ets; 1.
 DR PRINTS: PR00454; ETSDOMAIN.
 DR SMART: SM00413; ETS; 1.
 DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
 DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
 DR PROSITE: PS50061; ETS_DOMAIN_3; 1.
 KW Nuclear protein; Transcription regulation; Activator; DNA-binding.
 FT DOMAIN 75 80 POLY-ASP.
 FT DNA_BIND 208 290 ETS-DOMAIN.
 SQ SEQUENCE 612 AA; 66221 MM; 442F4C85142B31F0 CRC64;

Query Match 13.6%; Score 76.5; DB 1; Length 612;
 Best Local Similarity 24.7%; Pred. No. 3.9;
 Matches 38; Conservative 22; Mismatches 37; Indels 57; Gaps 9;

QY 2 QPGVYLS-QPAGIPGTSSSKOLFSLRHVVO-----QPSGNEKQVTT 42
 DB 363 QPSEVLEFVQPSQAPVPT---QLFRVHVVOVQVAPBEATIASTMQDEANSSVPSIRT 419
 QY 43 ISHSSTLTIT-----QKCGKTMPEVNTITPTSPASIIKQITLPQNKI 77
 DB 420 IQASTQVYVYVSPNQDLHYTVPLTVIASIDPSSGAGSKFTLITPSSQ-PMTVLK 477
 QY 78 QITL-----PQNKILSLQASPTQKNRI-KENVTS 105
 DB 478 ENVMQSQKRGSP--SIVLSPTQVQVLTFSNVGS 509

RESULT 6

YNR6_YEAST
 ID YNR6_YEAST STANDARD; PRT; 636 AA.
 AC P53882;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Hypothetical 67.4 kDa protein in RPS3-PSD1 intergenic region.
 GN YNL176C OR N1661.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycas.
 OX NCBI_Taxid=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Obermaier B., Piravandi E., Rinke M., Domdey H.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z71452; CAAG6068.1; -.
 DR SGD: S0005120; YNL176C.
 DR Hypothetical protein; Transmembrane.
 FT DOMAIN 94 98 POLY-SER.

```

FT DOMAIN 131 135 POLY-SER.
FT DOMAIN 301 309 POLY-SER.
FT DOMAIN 317 322 POLY-SER.
FT TRANSMEM 457 477 POTENTIAL.
SQ SEQUENCE 636 AA; 67356 MW; 4027EF48DEA5E287 CRC64;

Query Match 13.6%; Score 76.5; DB 1; Length 636;
Best Local Similarity 30.3%; Pred. No. 4.1;
Matches 23; Conservative 12; Mismatches 34; Indels 7; Gaps 2;

OY 7 LSOPAGIPTGSSSKQLFLFHVYQPSGNGEKQVTTISHSSTLTIOKGGQKTPVNTIIP 66
Db 296 ISVPTSSSVSSSSKVPKPS-----NRPSSSSSDDTTSAVSTYVQSLQSTTS--SSIP 348

OY 67 TSOEPPASILKQITLP 82
Db 349 TTOPPSTSTISTSPIP 364

RESULT 7
ELF1_HUMAN STANDARD: PRT; 619 AA.
ID P32519; O9UDE1.
AC 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ets-related transcription factor Elf-1 (E74-like factor 1).
GN ELF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE-92407982; PubMed-1527846;
RX Leiden J.M., Wang C.Y., Petryniak B., Markovitz D.M., Nabel G.J.,
RA Thompson C.B.;
RA "A novel Ets-related transcription factor, Elf-1, binds to human
RT immunodeficiency virus type 2 regulatory elements that are required
RL for inducible trans activation in T cells.";
RL J. Virol. 66:5890-5897(1992).
RN [2]
RP SEQUENCE OF 204-289 FROM N.A.
RX MEDLINE-92186836; PubMed-1545787;
RA Thompson C.B., Wang C.Y., Ho I.C., Bohjanen P.R., Petryniak B.,
RA June C.H., Miesfeldt S., Zhang L., Nabel G.J., Karpinski B.;
RT "Cis-acting sequences required for inducible interleukin-2 enhancer
RL function bind a novel Ets-related protein, Elf-1.";
RL Mol. Cell. Biol. 12:1043-1053(1992).
RN [3]
RP BINDING TO RB.
RX MEDLINE-93262492; PubMed-8493578;
RA Wang C.Y., Petryniak B., Thompson C.B., Kaelin W.G., Leiden J.M.;
RT "Regulation of the Ets-related transcription factor Elf-1 by binding
RL to the retinoblastoma protein.";
RL Science 260:1330-1335(1993).
CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT APPEARS TO BE REQUIRED FOR THE
CC 7-CELL-RECEPTOR-MEDIATED TRANS ACTIVATION OF HIV-2 GENE
CC EXPRESSION. BINDS SPECIFICALLY TO TWO PURINE-RICH MOTIFS IN THE
CC HIV-2 ENHANCER. ELF-1 BINDS TO THE UNDERPHOSPHORYLATED FORM OF RB.
CC MAY INTERACT WITH OTHER TRANSCRIPTION FACTORS IN ORDER TO REGULATE
CC SPECIFIC GENES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; M82882; -, NOT_ANNOTATED_CDS.
DR PIR; A43361; A43361.
DR HSSP; P28324; 1BC8.
DR TRANSEFAC; T01113; -.
DR Genew; HGNC:3316; ELF1.
DR MIM; 189733; -.
DR InterPro; IPR000418; Ets.
DR InterPro; IPR002341; HSF_ETS.
DR Pfam; PF00178; Ets; 1.
DR PRINTS; PR00454; ETSDOMAIN.
DR SMART; SM00413; ETS; 1.
DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
DR PROSITE; PS50061; ETS_DOMAIN_3; 1.
DR Nuclear protein; Transcription regulation; Activator; DNA-binding.
FT DOMAIN 75 80 POLY-ASP.
FT DNA_BIND 208 290 ETS-DOMAIN.
FT CONFLICT 283 283 Q -> G (IN REF. 2).
SQ SEQUENCE 619 AA; 67455 MW; AB0B41B2964A66EF CRC64;

Query Match 13.0%; Score 73; DB 1; Length 619;
Best Local Similarity 28.9%; Pred. No. 8.9;
Matches 37; Conservative 11; Mismatches 50; Indels 30; Gaps 6;

OY 1 AOPGVLVS--OPAGIPTGSSSKQLFLFHVYQO---PSGNGEKQVTTISHSSTLTIOKC 54
Db 362 AQPSEVLRLVPTQSGPYPR---QLFRVHVHPVQAVPEGG--EAARTSMODELTNSVQ 416

OY 55 GQKTPVNTIIPTSOEPASILKQITL-----PQKTLISQASPTQK- 96
Db 417 SIRTQAPQVQVNVVSPRNOQLHTVLTQVPLTVIASTPSAGTSGQKFLIQAIIPSSQ 476

OY 97 -NRKENV 103
Db 477 MTVLKENV 484

RESULT 8
SPA2_YEAST STANDARD: PRT; 1466 AA.
ID P23201.
AC 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE SPA2 protein.
GN SPA2 OR PEAL OR YLL021W OR L1209.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-S288C;
RX MEDLINE-91009481; PubMed-2211820;
RA Gehring S., Snyder M.;
RT "The SPA2 gene of Saccharomyces cerevisiae is important for
RL pheromone-induced morphogenesis and efficient mating.";
RL J. Cell Biol. 111:1451-1464(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-S288C;
RX STRAIN-S288C;
RA Purrelle B., Goffeau A.;
RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN PHEROMONE-INDUCED MORPHOGENESIS AND
CC -1- EFFICIENT MATING, PERHAPS AS A CYTOSKELETAL PROTEIN.
CC -1- SUBCELLULAR LOCATION: LOCALIZES A SHARP PATCH AT THE SHMOO TIP
CC (MATING PROJECTION) WHICH IS THE SITE OF POLARIZED CELL GROWTH.
CC -1- SIMILARITY: TO PROTEINS THAT FORM COILED-COIL STRUCTURES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch.)
 CC -----
 DR EMBL: X53731; CAA37763.1; -
 DR EMBL: X97560; CAA66170.1; -
 DR EMBL: Z73126; CAA97469.1; -
 DR PIR: A36426; A36426.
 DR SGD: S0003944; SPB2.
 KW Cell shape: Repeat; Coiled coil.
 FT DOMAIN 286 388 COILED COIL (POTENTIAL).
 FT 1087 818 1087 25 x 9 AA APPROXIMATE TANDEM REPEATS.
 FT REPEAT 818 825 1.
 FT REPEAT 825 834 2.
 FT REPEAT 835 843 3.
 FT REPEAT 860 868 4.
 FT REPEAT 875 883 5.
 FT REPEAT 884 892 6.
 FT REPEAT 893 901 7.
 FT REPEAT 902 910 8.
 FT REPEAT 911 919 9.
 FT REPEAT 920 928 10.
 FT REPEAT 929 937 11.
 FT REPEAT 938 946 12.
 FT REPEAT 947 953 13.
 FT REPEAT 954 961 14.
 FT REPEAT 962 970 15.
 FT REPEAT 971 979 16.
 FT REPEAT 980 988 17.
 FT REPEAT 989 997 18.
 FT REPEAT 998 1006 19.
 FT REPEAT 1007 1015 20.
 FT REPEAT 1036 1044 21.
 FT REPEAT 1045 1053 22.
 FT REPEAT 1054 1062 23.
 FT REPEAT 1072 1080 24.
 FT REPEAT 1081 1087 25.
 FT DOMAIN 1169 1189
 FT 1275 1302 COILED COIL (POTENTIAL).
 SQ DOMAIN 1466 AA; 163142 MW; 2EBB616152382C89 CRC64;
 Query Match 13.0%; Score 73; DB 1; Length 1466;
 Best Local Similarity 23.1%; Pred. No. 24;
 Matches 28; Conservative 27; Mismatches 32; Indels 34; Gaps 6;
 QY 5 PVL--SOPACIPGSS--SSKOLFSLF--HYVOQPSGNE-----KQVTTISHST 48
 DB 600 PIRPSSNGVPTTSRKPSGTGLFLMIDSSIAKNSHKEDNDKRYSPKAVTSASNSAS 659
 QY 49 LTIQKCGQKMP---VNTIIPTSOPPASILKQITLPGKILSLASPTOKRIKENVT 104
 DB 660 SNISELPKLTLPQAKIGVIPSE-----NOVPNIKIENTEEDNKRSDIT 705
 QY 105 s 105
 DB 706 N 706
 RESULT 9
 TIG_MYCGE
 ID TIG_MYCGE STANDARD; PRT; 444 AA.
 AC P47480;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Trigger factor (TF).
 DE TIG OR MG238.
 OS Mycoplasma genitalium.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;

RX MEDLINE=96026346; Pubmed=7569993;
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
 RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
 RA Nguyen D.T., Uitterlinden T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
 RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Luster T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
 RT "The minimal gene complement of Mycoplasma genitalium";
 RL Science 270:397-403(1995).
 RN [2]
 RP SEQUENCE OF 141-216 FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;
 RX MEDLINE=94075230; Pubmed=8253680;
 RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
 RA "A survey of the Mycoplasma genitalium genome by using random
 sequencing";
 RT J. Bacteriol. 175:7918-7930(1993).
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT. ACTS AS A CHAPERONE BY
 CC MAINTAINING THE NEWLY SYNTHESIZED PROTEIN IN AN OPEN CONFORMATION
 CC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PIPIASE FAMILY. TIG SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch.)
 CC -----
 DR EMBL: U39703; AAC71459.1; -
 DR EMBL: U01772; AAD10591.1; -
 DR TIG: MG238; -
 DR Interpro: IPR001179; FKBP_PPIase.
 DR Interpro: IPR005215; Trig_fac.
 DR Pfam: PF00254; FKBP; 1
 DR TIGRfam: TIGR00115; tig; 1.
 DR PROSITE: PS00453; FKBP_PPIASE_1; FALSE_NEG.
 DR PROSITE: PS00454; FKBP_PPIASE_2; FALSE_NEG.
 DR PROSITE: PS50059; FKBP_PPIASE_3; 1.
 DR Cell division: Chaperone; Isomerase; Rotamase; Complete proteome.
 FT DOMAIN 170 255
 FT 444 AA; 50931 MW; 99704A2BA4E23F47 CRC64;
 SQ SEQUENCE
 Query Match 12.9%; Score 72.5; DB 1; Length 444;
 Best Local Similarity 26.1%; Pred. No. 6.7;
 Matches 31; Conservative 18; Mismatches 31; Indels 39; Gaps 6;
 QY 10 PACIPTSSSKOLFSLFHYVOQPSGNEKQVTTISHSSTLTQKCGQKMPVNTIIPTSQ 69
 DB 267 PEOFOSTSLKAFKSYFHKIME---NKQETIIQENN---OKIRQFL-TWTKLIP--- 314
 QY 70 FPPASILKQITLPGKILSLASPTOKRI-----KENVT 104
 DB 315 FLPEALIK---LEANRLKLQSOAEQYKIPFEKLKLSASNTITLTELQDRNIKAKENVT 370
 RESULT 10
 REF1_HUMAN
 ID REF1_HUMAN STANDARD; PRT; 979 AA.
 AC P22670;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE MHC class II regulatory factor REF1 (REF) (Enhancer factor C) (EF-C).
 GN REF1.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91071581; Pubmed=2253877;

RA Reith W., Sanchez-Herrero C., Kober M., Silacci P., Berte C.,
RA Barras E., Mach B.;
RT "MHC class II regulatory factor RFX has a novel DNA-binding domain
and a functionally independent dimerization domain.";
RL Genes Dev. 4:1528-1540(1990).
RN [2]
RP IDENTITY BETWEEN RFX1 AND EF-C.
RX MEDLINE=94019311; PubMed=8413236;
RA Siegrist C.A., Durand B., Emery P., David E., Hearing P., Mach B.,
RA Reith W.;
RT "RFX1 is identical to enhancer factor C and functions as a
transactivator of the hepatitis B virus enhancer.";
RL Mol. Cell. Biol. 13:6375-6384(1993).
RN [3]
RP BINDING TO RPL30 PROMOTER.
RX MEDLINE=94040774; PubMed=8224874;
RA Safirany G., Perry R.P.;
RT "transcription factor RFX1 helps control the promoter of the mouse
ribosomal protein-encoding gene rpl30 by binding to its alpha
element.";
RL Gene 132:279-283(1993).
RN [4]
RP SHOWS THAT BLS IS NOT DUE TO RFX1.
RX MEDLINE=92375076; PubMed=1508204;
RA Sanchez-Herrero C., Reith W., Silacci P., Mach B.;
RT "The DNA-binding defect observed in major histocompatibility complex
class II regulatory mutants concerns only one member of a family of
RT complexes binding to the X boxes of class II promoters.";
RL Mol. Cell. Biol. 12:4076-4083(1992).
CC -1- FUNCTION: REGULATORY FACTOR ESSENTIAL FOR MHC CLASS II GENE
EXPRESSION. BINDS TO THE X BOXES OF MHC CLASS II GENES. ALSO BINDS
TO AN INVERTED REPEAT (ENH1) REQUIRED FOR HEPATITIS B VIRUS GENES
EXPRESSION AND TO THE MOST UPSTREAM ELEMENT (ALPHA) OF THE RPL30
PROMOTER.
CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE RFX FAMILY.
CC -1- SIMILARITY: BELONGS TO THE RFX FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X58964; GAA41730.1; -;
DR EMBL: A20498; GAA01506.1; -;
DR PIR: A35913; A35913.
DR TRANSFAC: T00909; -;
DR TRANSFAC: T01673; -;
DR Genew: HGNC:9982; RFX1.
DR MIM: 600006; -;
DR InterPro: IPR003150; RFX_DNA_binding.
DR Pfam: PF02257; RFX_DNA_binding.
KW DNA-binding; Transcription regulation; Activator; Nuclear protein.
FT DOMAIN 381 411 GLY-RICH.
FT DNA_BIND 438 528 EXPERIMENTALLY DEDUCED.
FT DOMAIN 920 936 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 744 979 NECESSARY FOR DIMERIZATION.
FT SEQUENCE 979 AA; 104728 MW; 556151F88C6AC9A2 CRC64;
SQ
Query Match 12.7%; Score 71.5; DB 1; Length 979;
Best Local Similarity 22.8%; Pred. No. 21;
Matches 34; Conservative 17; Mismatches 43; Indels 55; Gaps 6;

OY 86 --ILSLQA-----SPROKNIKRENTVS 105
DB 189 GGVSVLTWGHGQQVHSPPEQSPVOANSSS 217
RESULT 11
PODX_HUMAN
ID PODX_HUMAN STANDARD; PRT: 528 AA.
AC 000592;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Podocalyxin-like protein 1 precursor.
GN PODXL OR PCPL1 OR PCPLP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97332652; PubMed=9188463;
RA Kershaw D.B., Beck S.G., Wharram B.L., Wiggins J.E., Goyal M.,
RA Thomas P.E., Wiggins R.C.;
RT "Molecular cloning and characterization of human podocalyxin-like
protein. Orthologous relationship to rabbit PCPL1 and rat
podocalyxin.";
RL J. Biol. Chem. 272:15708-15714(1997).
CC -1- FUNCTION: Functions as an antiadhesin that maintains an open
filtration pathway between neighboring foot processes in the
podocyte by charge repulsion.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (podocyte).
CC -1- TISSUE SPECIFICITY: Glomerular epithelium cell (podocyte).
CC -1- PTM: Sialoglycoprotein.
CC -1- SIMILARITY: BELONGS TO THE PODOCALYXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U97519; AAB61574.1; -;
DR Genew: HGNC:9171; PODXL.
DR MIM: 602632; -;
KW Glycoprotein; Signal; Transmembrane; Polymorphism.
FT SIGNAL 1 22
FT CHAIN 23 528
FT DOMAIN 23 431
FT TRANSMEM 432 452
FT DOMAIN 453 528
FT DOMAIN 22 304
FT CARBOHYD 35 35
FT CARBOHYD 45 45
FT CARBOHYD 106 106
FT CARBOHYD 146 146
FT CARBOHYD 330 330
FT VARIANT 62 62
FT VARIANT 62 62
FT VARIANT 196 196
SQ SEQUENCE 528 AA; 55595 MW; 22A32B251F008D56 CRC64;
L -> S.
/FTID=VAR_012237.
Query Match 12.6%; Score 71; DB 1; Length 528;
Best Local Similarity 25.5%; Pred. No. 12;
Matches 27; Conservative 18; Mismatches 35; Indels 26; Gaps 5;

QY 59 -----MPVNTIIPTSQ-----FPFASLIKQITLPGNKILSLQASPT 94
 DB 248 QISSQMPASSTASQSEYQVPISPATALKRPTLPE-----TWSSSPT 289

RESULT 12
 NFE5_HUMAN STANDARD; PRT; 1531 AA.
 AC 094916; 095693; Q9UN18;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Nuclear factor of activated T cells 5 (T cell transcription factor
 NFAT5) (NF-AT5) (tonicity-responsive enhancer-binding protein) (Tone-
 binding protein) (ToneBP).
 GN NFAT5 OR TONEBP OR KIAA0827.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM C).
 RC TISSUE=Brain;
 RX MEDLINE=99156230; PubMed=10048485;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.,
 RT "Prediction of the coding sequences of unidentified human genes. XII.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 5:355-364(1998).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM A).
 RC TISSUE=Brain;
 RX MEDLINE=99307389; PubMed=10377394;
 RA Lopez-Rodriguez C., Aramburu J., Rakeman A.S., Rao A.;
 RT "NFAT5, a constitutively nuclear NFAT protein that does not cooperate
 RT with Fos and Jun.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:7214-7219(1999).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM A).
 RC TISSUE=Brain;
 RX MEDLINE=99162641; PubMed=10051678;
 RA Miyakawa H., Woo S.K., Dahl S.C., Handler J.S., Kwon H.M.;
 RT "Tonicity-responsive enhancer binding protein, a rel-like protein that
 RT stimulates transcription in response to hypertonicity.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:2538-2542(1999).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM B).
 RC TISSUE=Brain;
 RA Lopez-Rodriguez C., Aramburu J., Rakeman A.S., Copeland N.G.,
 RT Gilbert D.J., Thomas S., Distche C., Jenkins N.A., Rao A.;
 RT "NFAT5: The NF-AT family of transcription factors expands in a new
 RT direction.";
 RL Cold Spring Harb. Symp. Quant. Biol. 64:517-526(1999).
 RN [5]
 RP SEQUENCE OF 675-1531 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20029268; PubMed=10565538;
 RA Zuenke C., Kiehl R., Johannsmeier A., Grzeschik K.H., Schwinger E.;
 RT "Isolation and characterization of novel CAG repeat containing genes
 RT expressed in human brain.";
 RL DNA Seq. 10:1-6(1999).
 CC -1- FUNCTION: PLAYS A ROLE IN THE INDUCIBLE EXPRESSION OF GENES.
 CC REGULATES HYPERTONICITY-INDUCED CELLULAR ACCUMULATION OF
 CC OSMOLYTES.
 CC -1- SUBUNIT: DOES NOT BIND WITH FOS AND JUN TRANSCRIPTION FACTORS. BUT
 CC MIGHT BE CAPABLE OF FORMING STABLE DIMERS WITH DNA ELEMENTS.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: A, B AND C (SHOWN HERE); MAY BE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN SKELETAL MUSCLE, BRAIN,
 CC HEART AND PERIPHERAL BLOOD LEUCOCYTES. ALSO EXPRESSED IN PLACENTA,
 CC LUNG, LIVER, KIDNEY, PANCREAS, SPLEEN, THYMOS, PROSTATE, TESTIS,
 CC OVARY, SMALL INTESTINE AND COLON.

CC -1- SIMILARITY: BELONGS TO THE REL/DORSAL FAMILY.
 CC -1- CAUTION: REF.5 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 1165
 CC ONMAD DUE TO A FRAMESHIFT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.slb.ch).
 CC -----

DR EMBL: AB020634; BA74850.1; -
 DR EMBL: AF089624; AAD18136.1; -
 DR EMBL: AF134870; AAD38360.1; -
 DR EMBL: 297016; CAB09693.1; ALT_FRAME.
 DR EMBL: AF163836; AAD48441.1; -
 DR HSSP: 000653; 1A3Q.
 DR TRANSFAC: T04940; -
 DR GeneW: HGNC:7774; NFAT5.
 DR MIM: 604708; -
 DR InterPro: IPR002909; IPT_TIG.
 DR InterPro: IPR000451; NF_Rel_dor_fam.
 DR Pfam: PF01833; TIG: 1.
 DR SMART: SM00429; IPT_1.
 DR PROSITE: PS01204; REL_1; FALSE_NEG.
 DR PROSITE: PS0254; REL_2; 1.
 DR KW Transcription regulation; Activator; Nuclear protein; DNA-binding;
 KW Alternative splicing.
 FT DOMAIN 69 100 SER-RICH.
 FT DOMAIN 293 300 DNA-BINDING.
 FT DOMAIN 739 743 POLY-GLN.
 FT DOMAIN 879 888 POLY-GLN.
 FT DOMAIN 966 971 POLY-THR.
 FT DOMAIN 1248 1266 POLY-GLN.
 FT VARSPIC 1 76 MISSING (IN ISOFORM A).
 FT VARSPIC 1 47 MISSING (IN ISOFORM B).
 FT CONFLICT 1369 1369 E -> D (IN REF. 5).
 SQ SEQUENCE 1531 AA; 165764 MW; A68C6808BDA6F69E CRC64;

Query Match 12.6%; Score 71; DB 1; Length 1531;
 Best Local Similarity 29.3%; Pred. No. 41;
 Matches 29; Conservative 18; Mismatches 28; Indels 24; Gaps 6;

QY 1 ACPGCVLSQPA---GIPTSSSKQL-FSLFH--VVOQPSGNGEKQVYTRISHS-STLTIG 52
 DB 975 AFGTTFMFORSSGDSGETQAKQIQNSVFGTMVQHGSGDNQPVNLFSSSRKMSMSVQ 1034
 QY 53 KCGQKTPVNTIIPTSQFPFASLIKQITLPGNKILSLQ 91
 DB 1035 NSG-----TQQNGNGLFQQ---GNEMMSLQS 1057

RESULT 13
 DP3X_HAEIN STANDARD; PRT; 688 AA.
 AC P43746;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE DNA polymerase III subunit gamma/tau (EC 2.7.7.7).
 GN DNAX OR H11229.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA Mckenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,


```
FT DOMAIN 320 529 DMB.
FT DOMAIN 450 465 POLY-ALA.
FT CONFLICT 257 257 S -> A (IN REF. 2).
FT CONFLICT 292 292 R -> P (IN REF. 2).
SQ SEQUENCE 551 AA; 60417 MW; 0835EF88D9C1C980 CRC64;

Query Match 12.3%; Score 69; DB 1; Length 551;
Best Local Similarity 35.9%; Pred. No. 19;
Matches 23; Conservative 6; Mismatches 31; Indels 4; Gaps 2;

QY 8 SOPAGIPTGSSKQLFSLFHVQV-QPSGGNEKQVTTISHSTLTITKCGKQTMPVNTII 65
DB 222 SOPASTILAGSHSEGLLIQIASGPQGGQNGFTAQPSYHHNSTTT-WTGSRTAPYTPNL 279

QY 66 PTSQ 69
DB 280 PHHQ 283

RESULT 15
RGSK_MOUSE
ID RGSK_MOUSE STANDARD; PRT; 239 AA.
AC Q90ZB1; Q90ZB2; Q9CUV8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Regulator of G-protein signaling 20 (RGS20) (Regulator of G-protein
DE signaling 21).
GN RGS20 OR RGS21.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT ARG-32.
RC STRAIN=129/B6, and BALB/c;
RA Barker S.A., Wang J., Ross E.M.;
RT 'A mouse ortholog of RGS21.';
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 5-239 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Hippocampus;
RX MEDLINE=21085660; Pubmed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszew-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- FUNCTION: INHIBITS SIGNAL TRANSDUCTION BY INCREASING THE GTPASE
CC ACTIVITY OF G PROTEIN ALPHA SUBUNITS THEREBY DRIVING THEM INTO
CC THEIR INACTIVE GDP-BOUND FORM. BINDS SELECTELY TO G(12)-ALPHA AND
CC IS INHIBITED BY THE PHOSPHORYLATION AND PALMITOYLATION OF THE G-
CC PROTEIN (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -!- PTM: FATTY ACYLATED. HEAVILY PALMITOYLATED IN THE CYSTEINE STRING
CC MOTIF (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 RGS DOMAIN.
```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF191554; AAF05757.1; -;
DR EMBL; AF191552; AAF05756.1; -;
DR EMBL; AK013773; BAB28987.1; -;
DR HSSP; P49795; ICMZ.
DR MGD; MGI:1929866; Rgs20.
DR InterPro; IPR000342; Regl_Gprotein.
DR Pfam; PF00615; RGS; 1.
DR PRINTS; PR01301; RGS-PROTEIN.
DR ProDom; PD001580; Reg_of_prg; 1.
DR SMART; SM00313; RGS; 1.
DR PROSITE; PS50132; RGS; 1.
DR Signal transduction inhibitor; Membrane; Lipoprotein; Palmitate;
KW Polymorphism.
FT DOMAIN 59 71 POLY-CYS.
FT DOMAIN 113 229 RGS.
FT VARIANT 32 32 M -> R (IN BALB/C).
SQ SEQUENCE 239 AA; 26986 MW; F383923163A44D18 CRC64;

Query Match 12.2%; Score 68.5; DB 1; Length 239;
Best Local Similarity 21.1%; Pred. No. 8;
Matches 28; Conservative 23; Mismatches 53; Indels 29; Gaps 4;

QY 1 AQGPVLSQPAIGTSSSKQLFSLFHVQVQPSGGNEKQVTTISHS----- 46
DB 12 ASPSASPADP-GLPEGSEITEM-----RMQMGSGSETQGPAPSQQGGSGNACCCFWCC. 65
QY 47 -----SPLTI-----QKCGKQTMPVNTIIITSPPPASILKQITLPGNKILSLQASTQKN 97
DB 66 CCTCSCLTVRNOEDQRPQRASHEIRTDIPACESTPTLEEVCWAQSFNDLMVTTPAGRN 125
QY 98 RIKENVTSFRDE 110
DB 126 AFREFLRTFSEE 138

Search completed: February 16, 2003, 21:56:27
Job time : 6.11991 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:31:01 ; Search time 7.72591 Seconds
(without alignments)
2933.659 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_552
Perfect score: 562
Sequence: 1 AQP6PVLSPAGIPTGSSK.....ASPTQKNRIKENTVSCFRDE 110

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phage.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriap.*
 - 17: sp_archaeap.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98.5	17.5	662	11	Q91WW6 mus musculus
2	84.5	15.0	624	4	Q96JG5
3	81.5	14.5	643	3	Q94233
4	80.5	14.3	410	3	Q9UNK3
5	79.5	14.1	3726	5	Q9VFL1
6	78.5	14.0	506	5	Q8WQ86
7	78.5	14.0	612	5	Q8T0A8
8	78.5	14.0	622	5	Q9W4F6
9	78	13.9	346	16	Q9WVE1
10	77.5	13.8	545	2	Q52188
11	77.5	13.8	545	2	Q52193
12	77.5	13.8	751	5	Q960R2
13	77	13.7	918	13	Q9DG67
14	76	13.5	403	3	Q42963
15	76	13.5	696	4	Q9H9R3
16	75.5	13.4	754	16	Q8XTG8

17	74.5	13.3	2454	5	Q8T2G3
18	74	13.2	524	11	Q9EQY1
19	74	13.2	579	11	Q9EQY2
20	74	13.2	615	11	Q9EQY3
21	74	13.2	1844	5	Q22579
22	72.5	12.9	545	2	Q51643
23	72	12.8	327	4	Q9H683
24	72	12.8	370	11	Q9D3V7
25	72	12.8	1453	4	Q9Y6T1
26	72	12.8	1608	4	Q96RK0
27	72	12.8	2296	5	Q9VNE0
28	71.5	12.7	554	6	Q95L89
29	71.5	12.7	580	6	Q8WML4
30	71.5	12.7	642	3	Q60129
31	71	12.6	783	12	Q91331
32	71	12.6	1484	4	Q96QH3
33	71	12.6	2766	11	Q9QZR8
34	70.5	12.5	157	16	Q8REV0
35	70.5	12.5	397	10	Q94J09
36	70.5	12.5	487	10	Q9SY11
37	70.5	12.5	865	16	Q8YXD5
38	70	12.5	218	10	Q43395
39	70	12.5	410	4	Q95056
40	70	12.5	410	4	Q9UE14
41	70	12.5	424	12	Q99CX6
42	70	12.5	441	5	Q9XUS2
43	70	12.5	649	13	Q9DDC0
44	70	12.5	1072	4	Q9Y4G7
45	70	12.5	1484	11	P70098

ALIGNMENTS

RESULT 1

Q91WW6 PRELIMINARY; PRT; 662 AA.
ID Q91WW6
AC Q91WW6;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE TATA-binding protein associated factor TAFII135 (Fragment).
GN TAF4A.
DN TAF4A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE=PO BRAIN;
RA Metsis M., Brunkhorst A., Neuman T.;
RT "Cell Type Specific Expression of the TFIIID Component TAFII135 in the Nervous System."
RL Exp. Cell Res. 0:0-0(2001).
DR EMBL; AY038601; AAK94779.1;
DR MGD; MGI:2152346; Taf4a.
FT NON_TER
SQ SEQUENCE 662 AA; 71398 MW; 91A75F38CB0D0DA4 CRC64;

Query Match 17.5% Score 98.5; DB 11; Length 662;
Best Local Similarity 30.6%; Pred. No. 0.03;
Matches 36; Conservative 18; Mismatches 43; Indels 25; Gaps 7;

Qy	2	QPGLVS--QPAGIPTGSSSKQLFSLFHVQPSGGN---EKQVTIISHSSTLTITKQCG 56
Db	293	QP-PVISLTQPTQGVG---KQAPPTPLVIOQPKPGALIRPPQVTLTQTNPVALRQ--- 345
Qy	57	KTMVNTIIPF-----SQPPASILKQIYLPNGKIL---SLQASPTQKNRIKENTVSC 106
Db	346	---PHNRIMLTTPQOIQNLQNPVVPVVKPTVLPCTKALSTVSAQAAAAKKNLKEPGGGS 402
Qy	107	FRDE 110
		:

3

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Piannkoch C., Baldwin D.,
 RA Baller J.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Buhler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Posler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostlin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
 DR EMBL; AF003704; AAF55041.2; .
 DR HSSP; P20393; 1A6Y.
 DR FlyBase; FBgn0003862; trx.
 DR InterPro; IPR003889; FYrich_C.
 DR InterPro; IPR003888; FYrich_C.
 DR InterPro; IPR003616; PostSET.
 DR InterPro; IPR001214; SET.
 DR InterPro; IPR001965; Znf_PHD.
 DR InterPro; IPR001841; Znf_ring.
 DR Pfam; PF00628; PHD; 3.
 DR Pfam; PF00856; SET; 1.
 DR SMART; SM00542; FYRIC; 1.
 DR SMART; SM00541; FYRN; 1.
 DR SMART; SM00249; PHD; 4.
 DR SMART; SM00508; PostSET; 1.
 DR SMART; SM00184; RING; 3.
 DR SMART; SM00317; SET; 1.
 DR PROSITE; PS50280; SET; 1.
 SQ SEQUENCE 3726 AA; 40095 MW; E3DDB8F062BD7796 CRC64;
 Query Match 14.1%; Score 79.5; DB 5; Length 3726;
 Best Local Similarity 18.2%; Pred. No. 26;
 Matches 30; Conservative 24; Mismatches 50; Indels 61; Gaps 5;
 QY 3 PGPVLSQAPAGIPGSSSKQL--PSLFHVHVQPSGGNGKQV----- 40
 DB 2953 PKVTISQQR-IPATQTOQQQQAQMIHIPOQQPQQLOQQQVQVSMPIITLAEAPVQSQ 3011
 QY 41 -----TTTSHSTLTIOKCGKGTMPVNTIIPSTQPPASILKQITLPGNKI 86
 DB 3012 FVMEPQAEQOELANRVQHFSTSSSSSSSSSLFTNVNVMQQAQPSSTTSSTRPNRV 3071
 QY 87 LSLQ-----ASPTQKNRIKE-----NVTSCF 107
 DB 3072 LPMQORQEPAPLSNECPVSSPPPKPVEQPIIHQMTIASVSKY 3116

RESULT 6
 Q8WQ86 PRELIMINARY; PRT; 506 AA.
 AC Q8WQ86;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Microtubule-associated protein EBI.
 GN EBI.
 OS Dictyostellium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AX2;
 RA Rehberg M., Graef R.;
 RT "DeEBI is a permanent centrosomal resident in Dictyostellium and
 RT required for proper spindle formation.";
 RL Mol. Biol. Cell 128:309a-309a(2001).
 DR EMBL; AJ426053; CAD19801.1; .
 DR InterPro; IPR001715; Calponin-like.
 DR InterPro; IPR004953; Calponin-like.
 DR Pfam; PF00307; CH; 1.
 DR Pfam; PF03271; EBI; 1.
 DR PROSITE; PS50021; CH; 1.
 SQ SEQUENCE 506 AA; 56983 MW; F20BC182BAAA639C CRC64;
 Query Match 14.0%; Score 78.5; DB 5; Length 506;
 Best Local Similarity 32.0%; Pred. No. 3.3;
 Matches 32; Conservative 11; Mismatches 40; Indels 17; Gaps 4;
 QY 1 AQPGP-VLSQAPAGIPTGSS-----SKOLFSLFHVHVQPSGGNGKQVTTTSHSTLTIOK 54
 DB 175 AKTPSSSVKPSKPTSPSSISKVSKPTSKPTKPT-----PTTSTSTTTTSTP 228
 QY 55 GQKTPVNTIIPSTQPPASILKQITLPGNKILSLQASPT 94
 DB 229 PSTPKPTNPTIPSTGKPT-----LTQPTFKPTPKSVSPT 263
 RESULT 7
 Q8TOA8 PRELIMINARY; PRT; 612 AA.
 AC Q8TOA8;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE LD15062p.
 GN CG4088.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Colniker S.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY069439; AAL39584.1; .
 SQ SEQUENCE 612 AA; 65734 MW; E64C523CA2EA0623 CRC64;
 Query Match 14.0%; Score 78.5; DB 5; Length 612;
 Best Local Similarity 27.5%; Pred. No. 4.1;
 Matches 28; Conservative 19; Mismatches 40; Indels 15; Gaps 5;
 QY 5 PVLSPAPAGIPGSSSKQ--LFSLFHVHVQPSGGNGKQVTTTSHSTLTIOKCGKGTMPVN 62
 DB 499 PLKINVLGSGQGNKMKRAIPNSASSIOHENG-----VTTIVPASSLA---ASNQTAAMN 550

Qy 63 TIIP-TSQPPASILKQITLPGNKIISLQ-----ASPTQNRI 99
 | | | : | : | : | : | : | : | : | : | : | : | : | : | :
Dd 551 AIASTVTITPSSVGKKVTPNPKFILLKPAKFVGPASANEV 592

RESULT 8

Q9W4F6	Q9W4F6	PRELIMINARY;	PRT;	622 AA.
ID	Q9W4F6			
AC	Q9W4F6;			
DT	01-MAY-2000 (TrEMBLrel. 13, Created)			
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)			
DT	01-MAY-2000 (TrEMBLrel. 13, Last annotation update)			
DE	CS4068 protein.			
DE	CS4068			
GN	CS4068			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RP	STRAIN=BERKELEY;			
RC	MDLINE=20196006; PubMed=10731132;			
RX	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Ananidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,			
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,			
RA	Arrill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,			
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,			
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottiler P.,			
RA	Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Cawley S., Dahlike C., Davidson L.B., Davies P.,			
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.			
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.			
RA	Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,			
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,			
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibeguwa C.,			
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.			
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,			
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,			
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,			
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,			
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,			
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,			
RA	Palazzolo M., Pittman G.S., Pan S., Poilard J., Puri V., Reese M.G.,			
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,			
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,			
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,			
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,			
RA	Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,			
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,			
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,			
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu S., Smith H.O.,			
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;			
RT	"The genome sequence of Drosophila melanogaster.";			
RT	Science 287:2185-2195(2000).			
RL	EMBL; AE003432; AAF45997.1; -			
DR	FlyBase; Fgnd029738; CG4068;			
SQ	SEQUENCE 622 AA; 66817 MW; F6089D081D8D81A CRC64;			

Query Match 14.0%; Score 78.5; DB 5; Length 622;
Best Local Similarity 27.5%; Pred. NO. 4.2;
Matches 28; Conservative 19; Mismatches 40; Indels 15; Gaps 5;

QY 5 PVLSPAGIPTGSSKQ--LFSLFHWVQQPSGGNEKQVTTISHSSTLTIQCGOKTMPVN 62
|-| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| :| :|
Db 509 PLKINVLGSGSQGSNKMRAIFNSASSIOHENG-----VTTVIPASSLA---ASNQTAAAMN 560

QY 63 TIIP-TSQPPASILKQITLPGNKIISLQ-----ASPTQNRI 99
| | | : | : | : | : | : | : | :
Db 561 AIASTVTITPSSVGKKVTPNPKFILLKPAKFVGPASANEV 602

RESULT 9

Q9RYEI	PRELIMINARY;	PRT;	346 AA.
ID	Q9RYEI		
AC	Q9RYEI;		
DT	01-MAY-2000 (TrEMBLrel. 13, Created)		
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)		
DE	Hypothetical protein DR0008.		
DE	Hypothetical protein DR0008.		
GN	DR0008.		
OS	Deinococcus radiodurans.		
OC	Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;		
OC	Deinococcaceae; Deinococcus.		
OX	NCBI_TaxID=1299;		
RN	[1]		
SEQUENCE	FROM N.A.		
RP	STRAIN=RI;		
RC			
RX	MEDLINE=20036896; PubMed=10567266;		
RA	White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,		
RA	Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,		
RA	Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,		
RA	Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,		
RA	Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,		
RA	Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,		
RA	Fraser C.M.;		
RT	"genome sequence of the radioresistant bacterium Deinococcus		
RT	radiodurans RI";		
RL	Science 286:1571-1577(1999).		
DR	EMBL; AE001864; AAF09603.1; -.		
DR	TIGR; DR0008; -.		
KW	Hypothetical protein; Complete proteome.		
SK	SEQUENCE 346 AA; 37095 MW; 1E348C53CC4D2BA8 CRC64;		

Query Match 13.9%; Score 78; DB 16; Length 346;
Best Local Similarity 22.0%; Pred. No. 2.4;

[illegible]

RESULT 10

Q52188	PRELIMINARY;	PRT;	545 AA.
ID	Q52188		
AC	Q52188;		
DT	01-NOV-1996 (Tremblrel. 01, Created)		
DT	01-NOV-1996 (Tremblrel. 01, Last sequence update)		
DT	01-MAR-2002 (Tremblrel. 20, Last annotation update)		
DE	TaC precursor.		
DE	TRAC.		
GN	Enterococcus faecalis (Streptococcus faecalis).		
OS	Enterococcus faecalis (Streptococcus faecalis).		
OG	Plasmid pPDL.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;		
OC	Enterococcaceae; Enterococcus.		
OX	NCBI_TaxID=1351;		
RN	[1]		

[1]
RN SEQUENCE FROM N.A.
RP MEDLINE-9603293; PubMed-7559344;
RX Nakayama J., Yoshida K., Kobayashi H., Isogai A., Ciewell D.,
RY Suzuki A.;
RA "Cloning and characterization of a region of enterococcus faecalis
RT plasmid pp1 encoding pheromone inhibitor (ipd), pheromone sensitivity
RT (trac), and pheromone shutdown (traB) genes";

```

RL J. Bacteriol. 177:5567-5573(1995).
DR EMBL: D28859; BAA06007.1; -.
DR InterPro: IPR000914; SBP_bac_5.
DR Pfam: PF00496; SBP_bac_5; 2.
DR PROSITE; PS01040; SBP_BACTERIAL_5; UNKNOWN_1.
KW Plasmid; Signal.
FT SIGNAL 1 21 POTENTIAL.
SQ SEQUENCE 545 AA; 60780 MW; 0338A2D15E091C73 CRC64;

Query Match 13.8%; Score 77.5; DB 2; Length 545;
Best Local Similarity 29.5%; Pred. No. 4.6;
Matches 26; Conservative 13; Mismatches 28; Indels 21; Gaps 4;

Qy 15 TGSSSKOLFSLFHVVOQPSGNEKQVTTISHSSTLTITKCGKQTMPVNTIPTSQFPPAS 74
Db 135 TASPVELFSAIKNAKEIASG--KQV-----KDTLAVKSIQKTLTELVETPYF---- 183

Qy 75 ILKQITLPGNKILSLQA-SPTQKNRIKE 101
Db 184 -----TDLISLTAYYPVQKAIKE 202

RESULT 11
Q52193 PRELIMINARY; PRT; 545 AA.
ID AC Q52193
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE TRAC.
GN Enterococcus faecalis (Streptococcus faecalis).
OS Enterococcus faecalis (Streptococcus faecalis).
OC Plasmid pPDI.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN 1
RP SEQUENCE FROM N.A.
RX MEDLINE-96032393; PubMed-7559344;
RA Nakayama J., Yoshida K., Kobayashi H., Isogai A., Clewell D.,
RA Suzuki A.;
RT "Cloning and characterization of a region of enterococcus faecalis
RT plasmid pPDI encoding pheromone inhibitor (ipd), pheromone sensitivity
RT (trac), and pheromone shutdown (trab) genes.";
RL J. Bacteriol. 177:5567-5573(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-96032394; PubMed-7559345;
RA Fujimoto S., Tomita H., Wakamatsu E., Tanimoto K., Ike Y.;
RT "Physical Mapping of the Conjugative Bacteriocin Plasmid pPDI of
RT Enterococcus faecalis and Identification of the Determinant Related to
RT the Pheromone Response.";
RL J. Bacteriol. 177:5574-5581(1995).
DR EMBL: D78016; BAA11195.1; -.
DR InterPro: IPR000914; SBP_bac_5.
DR Pfam: PF00496; SBP_bac_5; 2.
DR PROSITE; PS01040; SBP_BACTERIAL_5; UNKNOWN_1.
KW Plasmid.
SQ SEQUENCE 545 AA; 60747 MW; 2277F1287C9D8291 CRC64;

Query Match 13.8%; Score 77.5; DB 2; Length 545;
Best Local Similarity 29.5%; Pred. No. 4.6;
Matches 26; Conservative 13; Mismatches 28; Indels 21; Gaps 4;

Qy 15 TGSSSKOLFSLFHVVOQPSGNEKQVTTISHSSTLTITKCGKQTMPVNTIPTSQFPPAS 74
Db 135 TASPVELFSAIKNAKEIASG--KQV-----KDTLAVKSIQKTLTELVETPYF---- 183

Qy 75 ILKQITLPGNKILSLQA-SPTQKNRIKE 101
Db 184 -----TDLISLTAYYPVQKAIKE 202

```

```

RESULT 12
Q960R2 PRELIMINARY; PRT; 751 AA.
ID AC Q960R2
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE LB39445p.
GN TRX OR CG8651.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY051904; AAK93328.1; -.
DR FlyBase; FBgn0003862; trx.
DR InterPro: IPR001214; SET.
DR Pfam: PF00856; SET; 1.
DR PROSITE; PS0280; SET; 1.
SQ SEQUENCE 751 AA; 83669 MW; C9E299C65A623C57 CRC64;

Query Match 13.8%; Score 77.5; DB 5; Length 751;
Best Local Similarity 19.2%; Pred. No. 6.7;
Matches 20; Conservative 22; Mismatches 37; Indels 25; Gaps 3;

Qy 29 VQPSGGNEKQVTT-ISHSSTLTITKCGKQTMPVNTIPTSQFPPASILKQITLPGNKIL 87
Db 38 VMEPQALEQQLANRVQHFSTSSSSSSSSSSSSSSSLPTNVNPMQOQAPSTSSSTRNRVL 97

Qy 88 SLQ-----ASPTQKNRIKE-----NVTSCF 107
Db 98 PMOQRPAPLNECPVVSPTPPKVEQPIHQMTSASVSKCY 141

RESULT 13
Q9DG67 PRELIMINARY; PRT; 918 AA.
ID AC Q9DG67
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Rad54b.
GN RAD54B.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN 1
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Yamaguchi-Iwai Y.;
RT "Rad54b, a Rad54-related gene product, is involved in recombinational
RT DNA repair.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF178529; AAG09308.1; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR000330; SNF2_N.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00176; SNF2_N; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
KW ATP-binding; Helicase.

```

```
SQ SEQUENCE 918 AA; 102878 MW; B4DB2AEA45006685 CRC64;
Query Match 13.7%; Score 77; DB 13; Length 918;
Best Local Similarity 27.4%; Pred. No. 9.6;
Matches 26; Conservative 14; Mismatches 49; Indels 6; Gaps 2;

QY 12 GPTGSSSKQLFSLFHVQVQSGGKQVTTSHSSTLTIQKQGTMPVNTIIPTSQPP 71
D 170 GRTGYKSKELDSLEBQTLIMIGKEIEVGVISADDFSSGRGCFQAGIATHDTVPTAL-- 227
QY 72 PASILKQITLPGNKILSLQASPTQKNRIKENVTSC 106
D 228 PQTTMRQFKCP-----IKSACQSPSTKENILLNPQSC 258

RESULT 14
O42963 PRELIMINARY; PRT; 403 AA.
AC O42963; P78918;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE Putative nucleoporin C1967.15.
GN SPBC1967.15.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBJ_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Lyne M., Rajandream M.A., Barrell B.G., Devlin K., Churcher C.M.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PR745;
RX MEDLINE=98162722; PubMed=9501991;
RA Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;
RT "Identification of open reading frames in Schizosaccharomyces pombe
RT CDNAs.";
RL DNA Res. 4:363-369(1997).
CC - FUNCTION: NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND
CC TRANSLLOCATION OF THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.
CC - SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX (POTENTIAL).
CC - SIMILARITY: TO YEAST NUCLEOPORIN NUP57.
DR EMBL; AL021839; CAA17069.1; -.
DR EMBL; D89269; BAA13930.1; ALT INIT.
KW Hypothetical protein; Nuclear protein; Transport.
FT CONFLICT 41 41 F -> L (IN REF. 2).
FT CONFLICT 75 75 T -> P (IN REF. 2).
FT CONFLICT 242 242 I -> F (IN REF. 2).
SQ SEQUENCE 403 AA; 44484 MW; 17FA9FBE1FE55C98 CRC64;

Query Match 13.5%; Score 76; DB 3; Length 403;
Best Local Similarity 27.0%; Pred. No. 4.7;
Matches 30; Conservative 14; Mismatches 49; Indels 18; Gaps 3;

QY 11 AGIPTGSSSKQLF--SLFHVQVQSGGKQVTTSHSSTLTIQKQGT----- 58
D 50 AGCGTSSSAPPFGNSIFGKTKQOQPTFSNTTTPAQSTVFGQNAASRTGNSNTQPLFS 109
QY 59 -----MPVNTIIPTSQPPASILKQITL-PGNKILSLQASPTQKNRIKENV 103
D 110 WSTVNNPTKPVDETNPATIPSSLLSSGSPNATVSNQAQGPAPPSVEQV 160

RESULT 15
Q9H8R3 PRELIMINARY; PRT; 696 AA.
AC Q9H8R3;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
```

```
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE OVARC1001306 protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBJ_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=OVARIAN CARCINOMA;
RA Isoqai T., Ota T., Hayaashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwavanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK023360; BAB14543.1; -.
DR InterPro; IPR001092; HLH_basic.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
FT NON_TER 696
SQ SEQUENCE 696 AA; 77150 MW; 9FA72151A4EA6A9E CRC64;

Query Match 13.5%; Score 76; DB 4; Length 696;
Best Local Similarity 23.0%; Pred. No. 8.9;
Matches 37; Conservative 23; Mismatches 41; Indels 60; Gaps 8;

QY 3 PGPVL-----SQAGIPTGSSSKQLFSLFHVQVQSGGK-----EKQVTT 42
D 5 PGSVMGIRLPAPSKPSETPPSTSSAFSYMNPVIOAVGSSAVNVITQAPSLSSGASF 64
QY 43 ISHSSTLTIQ-----KCGQKT-----MPVNT--IPTS-----QPP--- 71
D 65 VSOAGTTLTIRISPPPEQSPASKTSETKITYSGGQGPVGTASLPIQSGSFALLQPGQK 124
QY 72 --PASILKQITLPGNKILSLQASPTQKNRIKENVTSCFRDE 110
D 125 PVPSSILQHVA-----SLQMKRESQNPDKDKETNSIKRE 158

Search completed: February 16, 2003, 21:59:18
Job time : 12.7259 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:56:34 ; Search time 7.07238 Seconds
(without alignments)
1296.883 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_359

Perfect score: 1773

Sequence: 1 GLTVKVPVSPAPKVSQGP.....KAGVVLHSGVPTAATGTT 359

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 140259 seqs, 25548876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2.6/ptodata/2/pubpaa/US08_NEW_PUB.pap.*
- 2: /cgn2.6/ptodata/2/pubpaa/PCT_NEW_PUB.pap.*
- 3: /cgn2.6/ptodata/2/pubpaa/US06_NEW_PUB.pap.*
- 4: /cgn2.6/ptodata/2/pubpaa/US06_PUBCOMB.pap.*
- 5: /cgn2.6/ptodata/2/pubpaa/US07_NEW_PUB.pap.*
- 6: /cgn2.6/ptodata/2/pubpaa/US07_PUBCOMB.pap.*
- 7: /cgn2.6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pap.*
- 8: /cgn2.6/ptodata/2/pubpaa/US08_PUBCOMB.pap.*
- 9: /cgn2.6/ptodata/2/pubpaa/US09_NEW_PUB.pap.*
- 10: /cgn2.6/ptodata/2/pubpaa/US09_PUBCOMB.pap.*
- 11: /cgn2.6/ptodata/2/pubpaa/US10_NEW_PUB.pap.*
- 12: /cgn2.6/ptodata/2/pubpaa/US10_PUBCOMB.pap.*
- 13: /cgn2.6/ptodata/2/pubpaa/US60_NEW_PUB.pap.*
- 14: /cgn2.6/ptodata/2/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	491	27.7	1023	9	US-09-893-519A-14
2	167.5	9.4	5179	9	US-10-025-380-1068
3	167.5	9.4	5179	10	US-09-922-217-1068
4	167.5	9.4	5179	10	US-09-833-263-1068
5	162	9.1	1367	10	US-09-801-368-108
6	152.5	8.6	386	10	US-09-864-761-35720
7	151	8.5	1169	10	US-09-801-368-106
8	149.5	8.4	2665	10	US-09-864-761-34248
9	144	8.1	941	12	US-10-124-557-14
10	144	8.1	1022	12	US-10-124-557-84
11	144	8.1	1038	12	US-10-124-557-74
12	144	8.1	1049	12	US-10-124-557-58
13	144	8.1	1140	12	US-10-124-557-104
14	144	8.1	1270	12	US-10-124-557-44
15	144	8.1	1311	12	US-10-124-557-42
16	144	8.1	1313	12	US-10-124-557-142
17	144	8.1	1314	12	US-10-124-557-50
18	144	8.1	1320	12	US-10-124-557-46
19	144	8.1	1320	12	US-10-124-557-60

ALIGNMENTS

RESULT 1
US-09-893-519A-14
; Sequence 14, Application US/09893519A
; Publication No. US20030027243A1
; GENERAL INFORMATION:
; APPLICANT: ANADYS PHARMACEUTICALS, INC.
; APPLICANT: THOMPSON, Craig
; APPLICANT: MOORE, Jeffrey
; APPLICANT: BUURMAN, Ed T.
; APPLICANT: BRADLEY, John
; APPLICANT: DESILVA, Tamara
; APPLICANT: HARRIS, Sandra
; APPLICANT: KOMARNITSKY, Svetlana
; APPLICANT: MENDILLO, Marc
; APPLICANT: MOORE, Daniel
; APPLICANT: MCCOY, Melissa
; APPLICANT: SANDERSON, Karen
; APPLICANT: HAQ, Tariq
; APPLICANT: ZHU, Shuhao
; APPLICANT: LONG, Fan
; APPLICANT: DAVIDOV, Eugene
; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
; FILE REFERENCE: 0342/IG548-US2
; CURRENT APPLICATION NUMBER: US/09/893,519A
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,164
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/224,457
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 1023
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Corresponds to SEQ ID NO: 87
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: Human Genbank/CAA72189
; DATABASE ENTRY DATE: 1997-06-25
; RELEVANT RESIDUES: (1)...(1023)

Sequence 48, Appl
Sequence 40, Appl
Sequence 52, Appl
Sequence 2, Appl
Sequence 62, Appl
Sequence 110, Appl
Sequence 40, Appl
Sequence 195, Appl
Sequence 179, Appl
Sequence 353, Appl
Sequence 114, Appl
Sequence 34590, A
Sequence 40, Appl
Sequence 10, Appl
Sequence 6, Appl
Sequence 444, Appl
Sequence 104, Appl
Sequence 11, Appl
Sequence 14, Appl
Sequence 11, Appl
Sequence 16, Appl
Sequence 16, Appl
Sequence 24, Appl
Sequence 2, Appl

1354 12 US-10-124-557-48
1361 12 US-10-124-557-40
1363 12 US-10-124-557-52
1404 12 US-10-124-557-2
1404 12 US-10-124-557-62
1075 13 US-09-801-368-110
1481 10 US-09-371-900-40
1481 10 US-09-924-417-60
821 10 US-09-841-132-195
1776 10 US-09-841-132-179
583 10 US-09-841-132-353
1322 10 US-09-801-368-114
511 10 US-09-864-761-34590
1601 10 US-09-862-027-40
1056 9 US-10-161-510-10
1848 9 US-09-839-996-6
1770 10 US-09-841-132-444
1537 10 US-09-801-368-104
2586 10 US-09-905-129-11
2586 10 US-09-905-129-14
2586 10 US-09-991-630-11
2587 10 US-09-991-630-14
2587 10 US-09-905-129-16
2587 10 US-09-991-630-16
2589 10 US-09-991-630-24
2597 10 US-09-905-129-2

US-09-893-519A-14

Query Match 27.7%; Score 491; DB 9; Length 1023;
Best Local Similarity 38.4%; Pred. No. 6.5e-29;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

QY 13 PKVSSG-----PRLPAPQIVAVKAPNTTITQPPANLQPPGTVLKSNSSGGLM 61
DB 414 PTATSGIRATLPTVLAAPLPQP-----PONPTNIQ---NFQLPPGVMVLRSENGQLL 464
QY 62 LVSPQOTVTR-----AETTSNITSRRPAPVNPQVTKCTVPSNSSQLKKVAVTPVKKL 115
DB 465 MI-PQALAOQAQAHAPOTTMARPAITPSAPPVQISTVQAPGTPIIAR-QVTP----- 518
QY 116 AQIGTVTVTVTPKPSVQSVAVPTSVVTVTPGKPLNTVT--TLKPSSILGASS-----TPS 168
DB 519 -----TTIIKQV---SQOATTVQPSATLQKSPGVQQLVGGAAQTASLGATAVGTGPQ 571
QY 169 NE-PNLKAENSNAVINLSPHLENNVKKCNFLAMLIKACSGSOSPENGQVKKLVQOL 227
DB 572 RTVPGATTTSSAATE-----TMENVKCKNFKLSTLIKASSGKQSTETAANVKELVQNL 625
QY 228 LDKAEAEFEFTKLVKSSPOPHLVFPLKSVVALRQLLPNSOSFIOQCQQ-----TS 283
DB 626 LDKAEAEFEFTSLRELNSSPOPLVLFELKSLPALQLTPDLSAFLQSSQQQPPPPPTS 685
QY 284 SDMIATCTTTTTPVTVTVSSQSEKSIIVSGATAPRTVSV 327
DB 686 Q-----ATTALTAVLSSSVORTAGTAAATVTSALQPPVLSL 722

RESULT 2

US-10-025-380-1068
; Sequence 1068, Application US/10025380
; Publication No. US20020182191A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick Thomas S.
; APPLICANT: Carter, Darrick

; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C14

; CURRENT APPLICATION NUMBER: US/10/025.380

; CURRENT FILING DATE: 2001-12-19

; NUMBER OF SEQ ID NOS: 1129

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1068

; LENGTH: 5179

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-025-380-1068

Query Match 9.4%; Score 167.5; DB 9; Length 5179;
Best Local Similarity 22.08; Pred. No. 0.00096;
Matches 89; Conservative 39; Mismatches 180; Indels 97; Gaps 14;

QY 2 TLVTKVAPVSAPPKVSRRPAPQIVAVKAPNTTITQPPANLQPPGTVLKSNSSGGLM 61
DB 1607 TTTTTPPTTTPSPPTTPTP-PTSTTLPLPTTTPSPPTTTPPTT-----TTPSPPT 1661

QY 62 LVSPQOTVTRAE-----TSNITSRRPAPVNPQOT-----VKICTVPSNSSQL 103
DB 1662 TTTTSPPTTTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPSTT-- 1719
QY 104 IKKAVTPVKKLAQIGTIVTVTPKPSVQSVAVP-----TSVVTVTPGKPLNTVTVLKPS 159
DB 1720 -----TPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPSTP 1773
QY 160 SIGASSTP-----SNEPNLKAENSAVQIN--LSPTMLENVKCKNFKLAMLK 205
DB 1774 SITTTPPCVPLCNWGLDSCGPNFKPGDTELIGDVGCGWAANI----- 1821
QY 206 LACSSQSP-----EMGNVKKLVQOLLDKAEAEFEFTKLVKSSPOPHLVPLKKS 260
DB 1822 -SCRATMYDPVPIGQLGQTVVCDVSGLICKNEDQ-----KPG 1858
QY 261 VVALRQLLPNSOSFIOQCQQOTSSDMVIATCT--TTVTVSPVTVTVTVSSQSEKSIIVS 317
DB 1859 GVPMACFLNYEINVOCCECVTQPTTMTTMTTNTTENPPTTTPPTTTPPTTTPPTTPTGT 1918
QY 318 GATAPRTVSVQTLNPLNLAGPVGAK-----AGVVTLSHVGPTAATGGT 358
DB 1919 PTTTPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTGT 1962

RESULT 3

US-09-922-217-1068

; Sequence 1068, Application US/09922217

; Patent No. US2002007641A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13

; CURRENT APPLICATION NUMBER: US/09/922,217

; CURRENT FILING DATE: 2001-08-03

; NUMBER OF SEQ ID NOS: 1124

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1068

; LENGTH: 5179

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-922-217-1068

Query Match 9.4%; Score 167.5; DB 10; Length 5179;
Best Local Similarity 22.0%; Pred. No. 0.00096;
Matches 89; Conservative 39; Mismatches 180; Indels 97; Gaps 14;

QY 2 TLVTKVAPVSAPPKVSRRPAPQIVAVKAPNTTITQPPANLQPPGTVLKSNSSGGLM 61
DB 1607 TTTTTPPTTTPSPPTTPTP-PTSTTLPLPTTTPSPPTTTPPTT-----TTPSPPT 1661
QY 62 LVSPQOTVTRAE-----TSNITSRRPAPVNPQOT-----VKICTVPSNSSQL 103
DB 1662 TTTTSPPTTTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPSTT-- 1719
QY 104 IKKAVTPVKKLAQIGTIVTVTPKPSVQSVAVP-----TSVVTVTPGKPLNTVTVLKPS 159
DB 1720 -----TPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPSTP 1773
QY 160 SIGASSTP-----SNEPNLKAENSAVQIN--LSPTMLENVKCKNFKLAMLK 205

;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

;; FILE REFERENCE: Aemica-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; PRIOR FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 35720
;; LENGTH: 386
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AP000511.1
;; OTHER INFORMATION: EXPRESSED IN B474, SIGNAL = 1.7
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.3
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.7
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.4
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.1
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.5
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.5
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
;; OTHER INFORMATION: SWISSPROT HIT: P28968, EVALU 2.00e-05
US-09-864-761-35720

Query Match 8.6%; Score 152.5; DB 10; Length 386;
Best Local Similarity 22.0%; Pred. No. 0.00036;
Matches 85; Conservative 49; Mismatches 171; Indels 81; Gaps 13;
QY 2 TLTKVAPVSAPPKVSQGRPLPAPQ-----IVAVKAPNTTIOFPANLQLPPTVLK 54
DB 54 TTSSTVTVPPTTASTVSGSKNTACATSETTIAAITASEDTVS-----TQTSVIA 105
QY 55 SNSGP-----LMLVSPQOTVRAETSNITSRPAVPANPQVKICTVPNSSQLI- 104
DB 106 AESVPHATKTPTDITVASVSATVPKNTPTSVITSPS-----TAPNTASKWT 154

QY 105 --KKVAVTPVKKLAQIGTIVTTVPKPSVQSVAVPTSVVTVTPGKPLNTVTTLPKPSLG 162
DB 155 TASKATT--SUTISLPTTVFTTSK-----ITAGSEIPTASTTDSATTAISKAGTT 206
QY 163 ASSTPSNEPNLKAENSAVQINLSPTMLEN-----VKKCKNFLAMLIKACSGSQSP 216
DB 207 VESAPSTAPPTPAETTTASVPTTSTTGSENTGHTTVSSVPTTFATASESTGSETTR- 265
QY 217 GQNVKLVQQLDAKIEAEFEKRLVAKVELKSSPOPHLVFLKKSVALQLPNSQSFIQ 276
DB 266 -----ASTSATEVTMTTAMTGS-----TAVVSTKAPVTTTQSGFS 304
QY 277 QCV---QOTSSDMVIATCTTIVTTSPVTTTSSSSSEKSIIVSGATAPRTVSQVLNPL 333
DB 305 TATVMPAKTTTASVSTASTTLYQNTIDSVTKSVPTMDSTIASKSTTLSKIVSVPTAVFI 364
QY 334 AGPVGAKGVVTLHVSQVPTAATGGTT 359
DB 365 KAPETTTGSEITLASI-----ITSGTT 386

RESULT 7

US-09-801-368-106
; Sequence 106, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 106
; LENGTH: 1169
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-106

Query Match 8.5%; Score 151; DB 10; Length 1169;
Best Local Similarity 21.2%; Pred. No. 0.0021;
Matches 87; Conservative 69; Mismatches 137; Indels 118; Gaps 18;
QY 4 VTKVAPVSAPPKVSQGRPLPAPQIVAVKAPNTTIOF-PANLQLPPTVLKSNQPLML 62
DB 689 VTSSVWSTPITSSSESASVTIL---PSTITSEFKPMTK---TKVSISSSPTNL 740
QY 63 VSPQOTVTRAETSNITSRPAV-----PANPQVKICTVPNSSQLIKKAVTVP 112
DB 741 ITSYDTTSKDSITGSGSTSVSLISSISLPSYSASSEQIFHSSIVSSNGQATFSSTKV 800
QY 113 KK-----LAQIGTIVTTVPK-----PSSVQSVAVPTSVVTVTPGK 149
DB 801 SSSESSESHRTSPTTSSSESGIKSGVEIESTTSFSFHEITSTASTSVQISSQVTPSP 860
QY 150 LNTVTTLPSSLG-----ASSTPSNEPNLKAENSAVQINLSPTMLENKKCNFLAMLIK 205

Db 861 ISTVA---PRSTGLNSQTESTNSSKETMSSSENSASV-----893
Qy 206 LACSGSGSPENGQNNKLVLEOLLDAKIEAE-----FTR---KLXVELKSSQPHLVPLK 258
Db 894 MPSSATSPKGT-----KVTSDTSSGFSRDRTTVYRMTSETPTN-----EQ 936
Qy 259 KSVVALRQLPNSOSFIOQCQVQOQSSDMVIATCTTV-----TT-----SPVVTTVYSS 307
Db 937 TLTIVSSCESNCS-----NTVSSAVSTATITINGITTEYTTWCPLSATELTITVSK 989
Qy 308 SLS--EKSIIVGATAPRTVSVQTLNP--LAGPVGAKAGVVTLHVSVGPTAAT 355
Db 990 LESEBKTTLITVTSCEGVCSETASPAIVSTATATATVNDVTVTVYSTWSPOAT 1040

RESULT 8
US-09-864-761-34248
; Sequence 34248, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34248
; LENGTH: 2665
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL034555.2

; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 7.7
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 12
; OTHER INFORMATION: EST_HUMAN HIT: AUI17052.1, EVALUOE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P08640, EVALUOE 3.00e-10
US-09-864-761-34248
Query Match 8.4%; Score 149.5; DB 10; Length 2665;
Best Local Similarity 23.7%; Pred. No. 0.0086;
Matches 85; Conservative 50; Mismatches 161; Indels 63; Gaps 14;
Qy 9 PVSAPPKVSSGPRLPAPQIVAVKAPNTTIOFPANLQLPPGTVLKSKSGPLMLYS--PQ 66
Db 1841 PVTPP-----SDPSIPTPLPSVTA-----AKLSPVVASGGIPHSQPTKVTWITR 1887
Qy 67 QTVTRAETTSNITSRPAVPANPQTVKICTVPNSSSOLIKKVAVTP--VKLAQIGTTVVT 124
Db 1888 QEEPRASQTPS-----PALPDPDKASDVTSSSTL-RKILMDPKYVSATSVTSTSVTT 1939
Qy 125 TVPKPSSVQSVAVPTSVVVTGKPLNTVTTIKPSSSLGASSTPSNEPNLKAENSAVQIN 184
Db 1940 AIAEPVSAAPCLHEAPPPVDKPKLEKTA--PPTVNNSEIQASEVLVAADKERVAPV- 1996
Qy 185 LSPWLENVKKCKNFLAMLIKACSGSQSPENQGNKKLVQLLDKADAEFFTKLXVE 244
Db 1997 IAPKITSVISR-----MPVSIDLENSOKITLAKPAPQTLTGLVSA-----2036
Qy 245 LKSSQPQHLVPFLKKSVALROLPLNPSOSFIQCQVQOQSSDMVIATCTTTVTS- 302
Db 2037 LTGLVNVSLVP-----VNALGPKVGSVTTIKSLVSTPAGPVNVULGPNVLTGPVNLT 2091
Qy 303 TTVSSSQSEKSIIVSGATAPRTV--SVOTLNLAPGAVGAKAGVVTLHVSVGPTAATGGTT 359
Db 2092 TPVNAT-----VGTVNAAPGTVNAASAASAVNATASAVTGTAGAVTAASGGVTATTGTT 2144
RESULT 9
US-10-124-557-14
; Sequence 14, Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502

```
;
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 941 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-124-557-14

Query Match      8.1%; Score 144; DB 12; Length 941;
Best Local Similarity 20.3%; Pred. No. 0.0053;
Matches 75; Conservative 53; Mismatches 159; Indels 82; Gaps 12;

QY 8 APVSAPPKVSSGPRLPAPQIVAVKAPNT-----TTIOFPANLQLPQGVLI 53
DB 482 APTTPKEPAPTPKPEAPTPPKETAPTPKGTAPTTLKEPAPTPKKPAPKELAPT--- 538
QY 54 KNSGPLMLVSPQQTIVTRAETTSNITSRPAVPANPQTVKICTVPNSSQLIKKVAVTPVK 113
DB 539 --TKPTSTSDKPAPTPKGTAPTPKEPAPTPKGTAPTPKGTAPTLKEPAPTPK 596
QY 114 KLA-----QIGTVVTVTPKPSVQSVVA--VPTSVTVTPGKPLNTVTLKPSSLG 162
DB 597 KPAPKELAPTTTKGPTSTSDKPAPTPKETAPTPKGTAPTPKGTAPTPKGTAPTP 656
QY 163 ASSTP--SNEPNL--KAENSAAVQNLST--MLN-----VKKCKNFLAMLIKACSG 210
DB 657 EVSTPTTKETPTTIHKSPDESPKALNSPKPEPVPPTKPAATKPEMTTAA 716
QY 211 -----SOSPENGQNVKLVQLDIAEAEFTRKLYVELKSSQPQHLV 254
DB 717 KDKTTERDLRTTPTTAAAPKMTKETATTTKTESKITATT-----TQVSTTTQDIT 770
QY 255 PFLKKSVALRQLLNSQSFIOQCQQVQSSDMVIATCTTTTSPVTTTVSSSQSEKSI 314
DB 771 PF-KITTLKTTTLAPK-----VTTTKTITTEIMNKPEETAKPKDRA 812
QY 315 IVSGATAPR 323
DB 813 TNSKATTPK 821

RESULT 10
US-10-124-557-84
; Sequence 84, Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
```

```
;
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1022 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-10-124-557-84

Query Match      8.1%; Score 144; DB 12; Length 1022;
Best Local Similarity 20.3%; Pred. No. 0.006;
Matches 75; Conservative 53; Mismatches 159; Indels 82; Gaps 12;

QY 8 APVSAPPKVSSGPRLPAPQIVAVKAPNT-----TTIOFPANLQLPQGVLI 53
DB 563 APTTPKEPAPTPKPEAPTPPKETAPTPKGTAPTTLKEPAPTPKKPAPKELAPT--- 619
QY 54 KNSGPLMLVSPQQTIVTRAETTSNITSRPAVPANPQTVKICTVPNSSQLIKKVAVTPVK 113
DB 620 --TKPTSTSDKPAPTPKGTAPTPKEPAPTPKGTAPTPKGTAPTLKEPAPTPK 677
QY 114 KLA-----QIGTVVTVTPKPSVQSVVA--VPTSVTVTPGKPLNTVTLKPSSLG 162
DB 678 KPAPKELAPTTTKGPTSTSDKPAPTPKETAPTPKGTAPTPKGTAPTPKGTAPTP 737
QY 163 ASSTP--SNEPNL--KAENSAAVQNLST--MLN-----VKKCKNFLAMLIKACSG 210
DB 738 EVSTPTTKETPTTIHKSPDESPKALNSPKPEPVPPTKPAATKPEMTTAA 797
QY 211 -----SOSPENGQNVKLVQLDIAEAEFTRKLYVELKSSQPQHLV 254
DB 798 KDKTTERDLRTTPTTAAAPKMTKETATTTKTESKITATT-----TQVSTTTQDIT 851
QY 255 PFLKKSVALRQLLNSQSFIOQCQQVQSSDMVIATCTTTTSPVTTTVSSSQSEKSI 314
DB 852 PF-KITTLKTTTLAPK-----VTTTKTITTEIMNKPEETAKPKDRA 893
QY 315 IVSGATAPR 323
DB 894 TNSKATTPK 902

RESULT 11
US-10-124-557-74
; Sequence 74, Application US/10124557
; Patent No. US20020137894A1
```

GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 1038 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-10-124-557-74

Query Match 8.1%; Score 144; DB 12; Length 1038;
Best Local Similarity 20.3%; Pred. No. 0.0061;
Matches 75; Conservative 53; Mismatches 159; Indels 82; Gaps 12;
QY 8 APVSAPKVSQGLRLPAPQIVAVKAPNT-----TTIQPANLQLPPGVLI 53
DB 547 APTTPKEPATTPKEPATTPKGTAPTKGTAPTKLKEPATTPKPKAPKELAPTT--- 603
QY 54 KNSGGLMLVSPQQTIVTRAETTSNITSRPVAVPANQVTKICTVNPSSQLIKKVAIVPVK 113
DB 604 --TKPTSTSDKPAITPKGTAPTKKEPATTPKKEPATTPKGTAPTKLKEPATTPK 661
QY 114 KLA-----QIGTVTVTVTPKPSVQVA--VPTSVVTVTPGKPLNVTTLKPSLSG 162
DB 662 KPAPKELAPTTTKGPTSTSDKPAITPKGTAPTKKEPATTPKKEPATTPKTPPTTS 721
QY 163 ASSTP--SNEPNL--KAENSAVQINLSPT--MLEN-----VKCKKNFLAMLIKACSG 210
DB 722 EVSTPTTKPTTHKSPDSELSNEPTPKALENSPKPEPGVPTTKTPAATPEMTTIA 781
QY 211 -----SQSPMGONVKVLQELLDKAEAEFTKLYVELKSSQPQHLV 254
DB 782 KDKTTERDLRTPTTETTTAAAPKMKETATTTTEKTESKITATT-----TQVTSITQDIT 835

QY 255 PFLKSVVALRQLLPNSQSFIOQCVOQTSSDMVIATCTTTTTPSPVVTITVSSQSEKSI 314
DB 836 PF-KITTLKTTTLAPK-----VTTTKKTTITTEIMNKPEETAKPKDRA 877
QY 315 IVSGATAPR 323
DB 878 TNSKATTPK 886
RESULT 12
US-10-124-557-58
Sequence 58, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1049 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 58:
US-10-124-557-58

Query Match 8.1%; Score 144; DB 12; Length 1049;
Best Local Similarity 20.3%; Pred. No. 0.0062;
Matches 75; Conservative 53; Mismatches 159; Indels 82; Gaps 12;
QY 8 APVSAPKVSQGLRLPAPQIVAVKAPNT-----TTIQPANLQLPPGVLI 53
DB 590 APTTPKEPATTPKKEPATTPKGTAPTKGTAPTKLKEPATTPKPKAPKELAPTT--- 646
QY 54 KNSGGLMLVSPQQTIVTRAETTSNITSRPVAVPANQVTKICTVNPSSQLIKKVAIVPVK 113
DB 647 --TKPTSTSDKPAITPKGTAPTKKEPATTPKKEPATTPKGTAPTKLKEPATTPK 704

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 22:01:05 ; Search time 126.688 Seconds
(without alignments)
1827.005 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_359
Perfect score: 1773
Sequence: 1 GTLVTKVAPVAPPKVSPP.....KAGVYTLHSVGPTRAGGTT 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_Main:*

- 1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pcp.*
- 2: /cgn2_6/ptodata/2/paa/US06_COMB.pcp.*
- 3: /cgn2_6/ptodata/2/paa/US07_COMB.pcp.*
- 4: /cgn2_6/ptodata/2/paa/US08_COMB.pcp.*
- 5: /cgn2_6/ptodata/2/paa/US081_COMB.pcp.*
- 6: /cgn2_6/ptodata/2/paa/US082_COMB.pcp.*
- 7: /cgn2_6/ptodata/2/paa/US083_COMB.pcp.*
- 8: /cgn2_6/ptodata/2/paa/US084_COMB.pcp.*
- 9: /cgn2_6/ptodata/2/paa/US085_COMB.pcp.*
- 10: /cgn2_6/ptodata/2/paa/US086_COMB.pcp.*
- 11: /cgn2_6/ptodata/2/paa/US087_COMB.pcp.*
- 12: /cgn2_6/ptodata/2/paa/US088_COMB.pcp.*
- 13: /cgn2_6/ptodata/2/paa/US089_COMB.pcp.*
- 14: /cgn2_6/ptodata/2/paa/US090_COMB.pcp.*
- 15: /cgn2_6/ptodata/2/paa/US091_COMB.pcp.*
- 16: /cgn2_6/ptodata/2/paa/US092_COMB.pcp.*
- 17: /cgn2_6/ptodata/2/paa/US093_COMB.pcp.*
- 18: /cgn2_6/ptodata/2/paa/US094_COMB.pcp.*
- 19: /cgn2_6/ptodata/2/paa/US095_COMB.pcp.*
- 20: /cgn2_6/ptodata/2/paa/US096_COMB.pcp.*
- 21: /cgn2_6/ptodata/2/paa/US097_COMB.pcp.*
- 22: /cgn2_6/ptodata/2/paa/US098_COMB.pcp.*
- 23: /cgn2_6/ptodata/2/paa/US099_COMB.pcp.*
- 24: /cgn2_6/ptodata/2/paa/US100_COMB.pcp.*
- 25: /cgn2_6/ptodata/2/paa/US101_COMB.pcp.*
- 26: /cgn2_6/ptodata/2/paa/US102_COMB.pcp.*
- 27: /cgn2_6/ptodata/2/paa/US60_COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1773	100.0	801	1 PCT-US01-08631-39827	Sequence 39827, A
2	1773	100.0	843	1 PCT-US02-29964-410	Sequence 410, App
3	1773	100.0	852	21 US-09-763-909-2	Sequence 2, Appl1
4	1761.5	99.4	865	1 PCT-US02-25829-25	Sequence 25, Appl
5	1575	88.8	685	27 US-60-243-468-1271	Sequence 1271, Ap
6	491	27.7	737	1 PCT-US94-01114-16	Sequence 16, Appl

7	491	27.7	1023	1 PCT-US01-20592-14	Sequence 14, Appl
8	491	27.7	1023	1 PCT-US01-20592A-14	Sequence 14, Appl
9	491	27.7	1023	22 US-09-893-519A-14	Sequence 14, Appl
10	491	27.7	1051	1 PCT-US02-30474-1660	Sequence 1660, Ap
11	491	27.7	1051	27 US-60-324-631-1665	Sequence 1665, Ap
12	479	27.0	1083	1 PCT-US02-30474-3271	Sequence 3271, Ap
13	218	12.3	899	27 US-60-167-217-11401	Sequence 11401, A
14	218	12.3	921	27 US-60-173-464-9223	Sequence 9223, Ap
15	218	12.3	921	27 US-60-173-464-20611	Sequence 20611, A
16	216	12.2	921	1 PCT-US94-01114-2	Sequence 2, Appl1
17	216	12.2	921	4 US-08-013-412-2	Sequence 2, Appl1
18	216	12.2	921	20 US-09-614-150-11376	Sequence 11376, A
19	216	12.2	921	20 US-09-614-150-24957	Sequence 24957, A
20	216	12.2	921	27 US-60-191-637-11408	Sequence 11408, A
21	216	12.2	921	27 US-60-191-637-25058	Sequence 25058, A
22	216	12.2	921	27 US-60-191-681-8933	Sequence 8933, Ap
23	216	12.2	921	27 US-60-191-681-19739	Sequence 19739, A
24	182	10.3	1795	20 US-09-614-150-36210	Sequence 36210, A
25	182	10.3	1795	27 US-60-191-637-35824	Sequence 35824, A
26	182	10.3	1795	27 US-60-191-681-28036	Sequence 28036, A
27	175.5	9.9	708	1 PCT-US01-04098A-3624	Sequence 45515, A
28	175.5	9.9	709	1 PCT-US01-04098A-1656	Sequence 1656, Ap
29	175.5	9.9	709	20 US-09-629-469A-11876	Sequence 11876, A
30	175.5	9.9	709	20 US-09-629-469A-14832	Sequence 14832, A
31	175.5	9.9	709	20 US-09-611-526-3857	Sequence 1119, Ap
32	175.5	9.9	881	1 PCT-US02-09921-1119	Sequence 1119, Ap
33	175.5	9.9	881	20 US-09-840-746-20	Sequence 20, Appl
34	173.5	9.8	528	22 US-60-185-361-566	Sequence 566, App
35	168	9.5	1953	27 US-60-185-361-494	Sequence 494, App
36	168	9.5	2947	27 US-60-185-361-938	Sequence 938, App
37	168	9.5	2947	20 US-09-641-377-697	Sequence 697, App
38	167.5	9.4	692	20 US-09-641-377-686	Sequence 686, App
39	167.5	9.4	717	20 US-10-221-279-12349	Sequence 12349, A
40	167.5	9.4	4742	26 US-02-17382-218	Sequence 218, App
41	167.5	9.4	5179	19 US-09-538-092-1258	Sequence 1258, Ap
42	167.5	9.4	5179	20 US-09-509-448A-1068	Sequence 1068, Ap
43	167.5	9.4	5179	20 US-09-849-811-1068	Sequence 1068, Ap
44	167.5	9.4	5179	22 US-09-833-263-1068	
45	167.5	9.4	5179	22 US-09-833-263-1068	

ALIGNMENTS

RESULT 1
PCT-US01-08631-39827
; Sequence 39827, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 39827
; LENGTH: 801
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-08631-39827

Query Match	100.0%	Score 1773;	DB 1;	Length 801;		
Best Local Similarity	100.0%	Pred. No. 4.6e-139;				
Matches 359;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
Qy	1	GTLVTKVAPVAPPKVS	GGPRLP	APQIVAVKAPNTTTIQFPANLQLP	PCTVLIKSQPL	60
Db	1	GTLVTKVAPVAPPKVS	GGPRLP	APQIVAVKAPNTTTIQFPANLQLP	PCTVLIKSQPL	60

```
QY 61 MLVSPQQTVTRAEETTSNITSRPAVPANPQTVKICTVPNSSQLIKKAVTPVKKLAQIGT 120
Db 61 MLVSPQQTVTRAEETTSNITSRPAVPANPQTVKICTVPNSSQLIKKAVTPVKKLAQIGT 120
QY 121 TVTIVTPKPSVQSVAVPTSVVTVTPGKPLNTVTLKPSLIGASSTPSNEPNLKAENSAA 180
Db 121 TVTIVTPKPSVQSVAVPTSVVTVTPGKPLNTVTLKPSLIGASSTPSNEPNLKAENSAA 180
QY 181 VOINLSPMTLENVKKCNFLAMLILKACSGSQSPENGQNVKKLVEQLDLDAKIEAEFEFTRK 240
Db 181 VOINLSPMTLENVKKCNFLAMLILKACSGSQSPENGQNVKKLVEQLDLDAKIEAEFEFTRK 240
QY 241 LYVELKSSPQPHLVFELKKSVALRQLLPNSQSFTQQCVQQTSSDMVIATCTTTVTTSVP 300
Db 241 LYVELKSSPQPHLVFELKKSVALRQLLPNSQSFTQQCVQQTSSDMVIATCTTTVTTSVP 300
QY 301 VTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTVLHSGVPTAATGGTT 359
Db 301 VTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTVLHSGVPTAATGGTT 359
```

RESULT 2

```
PCT-US02-29964-410
; Sequence 410, Application PC/TUS0229964
; GENERAL INFORMATION:
```

```
; APPLICANT: Tang, Y. Tom
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhang, Jie
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Xue, Aigong J.
; APPLICANT: Wang, Dunrui
; APPLICANT: Ghosh, Malabika
; APPLICANT: Asundi, Vinod
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle W.
; APPLICANT: Weng, Gezhi
; APPLICANT: Haley-Vicente, Dana
; APPLICANT: Drmanac, Radolje T
; TITLE OF INVENTION: Novel Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 809ACIP PCT
; CURRENT APPLICATION NUMBER: PCT/US02/29964
; PRIOR FILING DATE: 2002-09-19
; PRIOR APPLICATION NUMBER: US 60/323,739
; PRIOR FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 992
; SOFTWARE: pc_flg-genes version 6.0
; SEQ ID NO 410
; LENGTH: 843
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-29964-410
```

```
Query Match 100.0%; Score 1773; DB 1; Length 843;
Best Local Similarity 100.0%; Pred. No. 5e-139;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTLVTKVAPVSAPPKVSQSGPRLPAPQIVAVKAPNTTTIOFPANLQLPPGTVLIKSNGL 60
Db 43 GTLVTKVAPVSAPPKVSQSGPRLPAPQIVAVKAPNTTTIOFPANLQLPPGTVLIKSNGL 102
QY 61 MLVSPQQTVTRAEETTSNITSRPAVPANPQTVKICTVPNSSQLIKKAVTPVKKLAQIGT 120
Db 103 MLVSPQQTVTRAEETTSNITSRPAVPANPQTVKICTVPNSSQLIKKAVTPVKKLAQIGT 162
QY 121 TVTIVTPKPSVQSVAVPTSVVTVTPGKPLNTVTLKPSLIGASSTPSNEPNLKAENSAA 180
Db 163 TVTIVTPKPSVQSVAVPTSVVTVTPGKPLNTVTLKPSLIGASSTPSNEPNLKAENSAA 222
QY 181 VOINLSPMTLENVKKCNFLAMLILKACSGSQSPENGQNVKKLVEQLDLDAKIEAEFEFTRK 240
Db 223 VOINLSPMTLENVKKCNFLAMLILKACSGSQSPENGQNVKKLVEQLDLDAKIEAEFEFTRK 282
QY 241 LYVELKSSPQPHLVFELKKSVALRQLLPNSQSFTQQCVQQTSSDMVIATCTTTVTTSVP 300
Db 283 LYVELKSSPQPHLVFELKKSVALRQLLPNSQSFTQQCVQQTSSDMVIATCTTTVTTSVP 342
QY 301 VTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTVLHSGVPTAATGGTT 359
Db 343 VTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTVLHSGVPTAATGGTT 401

RESULT 3
US-09-763-909-2
; Sequence 2, Application US/09763909
; GENERAL INFORMATION:
; APPLICANT: Dikstein, Rivka
; APPLICANT: Yamit-Hezi, Ayala
; TITLE OF INVENTION: A TRANSCRIPTION FACTOR TFIIID SUBUNIT,
; TITLE OF INVENTION: TAFII105, POLYPEPTIDES, DNA ENCODING THEREFOR AND
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS
; FILE REFERENCE: 13005/002001
; CURRENT APPLICATION NUMBER: US/09/763,909
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 852
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-763-909-2
```

```
Query Match 100.0%; Score 1773; DB 21; Length 852;
Best Local Similarity 100.0%; Pred. No. 5e-139;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTLVTKVAPVSAPPKVSQSGPRLPAPQIVAVKAPNTTTIOFPANLQLPPGTVLIKSNGL 60
Db 1 GTLVTKVAPVSAPPKVSQSGPRLPAPQIVAVKAPNTTTIOFPANLQLPPGTVLIKSNGL 60
QY 61 MLVSPQQTVTRAEETTSNITSRPAVPANPQTVKICTVPNSSQLIKKAVTPVKKLAQIGT 120
Db 61 MLVSPQQTVTRAEETTSNITSRPAVPANPQTVKICTVPNSSQLIKKAVTPVKKLAQIGT 120
QY 121 TVTIVTPKPSVQSVAVPTSVVTVTPGKPLNTVTLKPSLIGASSTPSNEPNLKAENSAA 180
Db 121 TVTIVTPKPSVQSVAVPTSVVTVTPGKPLNTVTLKPSLIGASSTPSNEPNLKAENSAA 180
QY 181 VOINLSPMTLENVKKCNFLAMLILKACSGSQSPENGQNVKKLVEQLDLDAKIEAEFEFTRK 240
Db 181 VOINLSPMTLENVKKCNFLAMLILKACSGSQSPENGQNVKKLVEQLDLDAKIEAEFEFTRK 240
QY 241 LYVELKSSPQPHLVFELKKSVALRQLLPNSQSFTQQCVQQTSSDMVIATCTTTVTTSVP 300
Db 241 LYVELKSSPQPHLVFELKKSVALRQLLPNSQSFTQQCVQQTSSDMVIATCTTTVTTSVP 300
```

QY 301 VTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVVTLSHVGPTAATGCTT 359
Db 301 VTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVVTLSHVGPTAATGCTT 359

RESULT 4

PCT-US02-25829-25

Sequence 25, Application PC/TUS0225829

GENERAL INFORMATION:

APPLICANT: INCYTE GENOMICS, INC.

APPLICANT: BAROSSO, Ines

APPLICANT: BAUGHN, Mariah R.

APPLICANT: BECHA, Shanya D.

APPLICANT: BLAKE, Julie J.

APPLICANT: BOROWSKY, Mark L.

APPLICANT: BURFORD, Neil

APPLICANT: DUGGAN, Brendan M.

APPLICANT: ELLIOTT, Vicki S.

APPLICANT: EMERLING, Brooke M.

APPLICANT: FORSYTHE, Ian J.

APPLICANT: GIETZEN, Kimberly J.

APPLICANT: GORVAD, Ann E.

APPLICANT: GRIFFIN, Jennifer A.

APPLICANT: HAFALIA, April J.A.

APPLICANT: HONCHELL, Cynthia D.

APPLICANT: ISON, Craig H.

APPLICANT: KHAN, Farrah A.

APPLICANT: LAL, Preeti G.

APPLICANT: LEE, Ernestine A.

APPLICANT: LEE, Sally

APPLICANT: LEE, Soo Yeun

APPLICANT: LI, Joana X.

APPLICANT: LU, Dyung Aina M.

APPLICANT: LU, Yan

APPLICANT: LEHR-MASON, Patricia M.

APPLICANT: NGUYEN, Dannel B.

APPLICANT: RAKUMAR, Jayalaxmi

APPLICANT: SPRAGUE, William W.

APPLICANT: TANG, Y. Tom

APPLICANT: THANGAVELU, Kavitha

APPLICANT: THORNTON, Michael

APPLICANT: TRAN, Uyen K.

APPLICANT: WALIA, Narinder K.

APPLICANT: WARREN, Bridget A.

APPLICANT: XU, Yuming

APPLICANT: YAO, Monique G.

APPLICANT: YUE, Henry

APPLICANT: YUE, Huibin

APPLICANT: ZEBARADIAN, Yeganeh

TITLE OF INVENTION: NUCLEIC ACID ASSOCIATED PROTEINS

FILE REFERENCE: PF-1146 PCT

CURRENT APPLICATION NUMBER: PCT/US02/25829

CURRENT FILING DATE: 2002-08-14

PRIOR APPLICATION NUMBER: US 60/313,111

PRIOR FILING DATE: 2001-08-17

PRIOR APPLICATION NUMBER: US 60/314,682

PRIOR FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: US 60/314,756

PRIOR FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: US 60/315,105

PRIOR FILING DATE: 2001-08-27

PRIOR APPLICATION NUMBER: US 60/316,751

PRIOR FILING DATE: 2001-08-31

PRIOR APPLICATION NUMBER: US 60/316,856

PRIOR FILING DATE: 2001-08-31

PRIOR APPLICATION NUMBER: US 60/328,185

PRIOR FILING DATE: 2001-10-05

NUMBER OF SEQ ID NOS: 66

SOFTWARE: PERL Program

SEQ ID NO 25

LENGTH: 865

TYPE: PRT

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: 4398735CD1
PCT-US02-25829-25

Query Match 99.4%; Score 1761.5; DB 1; Length 865;
Best Local Similarity 99.2%; Pred. No. 4.8e-138;
Matches 359; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 GTLVTKVAPVSAPPKVSSGPRLPAPQIVAVKAPNTTTIQFPANLQLPP---GTVLIKSNS 57
Db 67 GTLVTKVAPVSAPPKVSSGPRLPAPQIVAVKAPNTTTIQFPANLQLPPGMLGTVLKSNS 126
QY 58 GPLMLVSPQQTVTTRAETTSNITSRPAPVNPQTVKICTVPSNSQLIKKVAVTPVKKLAQ 117
Db 127 GPLMLVSPQQTVTTRAETTSNITSRPAPVNPQTVKICTVPSNSQLIKKVAVTPVKKLAQ 186
QY 118 IGTTVTTVPKPSVQSVAVPTSVVTVTPGKPLNTVTTLPKPSLSGASSTPSNEPNAEN 177
Db 187 IGTTVTTVPKPSVQSVAVPTSVVTVTPGKPLNTVTTLPKPSLSGASSTPSNEPNAEN 246
QY 178 SAAVOINLSPTMLENVKCKNFNLAKCSGSSQSPGEMGNVKKLVQELLDKIEAEF 237
Db 247 SAAVOINLSPTMLENVKCKNFNLAKCSGSSQSPGEMGNVKKLVQELLDKIEAEF 306
QY 238 TRKLYVELKSSQPPLVPLFKKSVVALRQLLPNSQSFIOQCVOQTSSDMVIATCTTTVT 297
Db 307 TRKLYVELKSSQPPLVPLFKKSVVALRQLLPNSQSFIOQCVOQTSSDMVIATCTTTVT 366
QY 298 SPVTTTSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVVTLSHVGPTAATGG 357
Db 367 SPVTTTSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVVTLSHVGPTAATGG 426
QY 358 TT 359
Db 427 TT 428

RESULT 5

US-60-243-468-1271

Sequence 1271, Application US/60243468

GENERAL INFORMATION:

APPLICANT: Beasley, Ellen

TITLE OF INVENTION: ISOLATED HUMAN DRUG TARGET PROTEINS,

NUCLEIC ACID MOLECULES ENCODING HUMAN DRUG TARGET PROTEINS,

TITLE OF INVENTION: AND USES THEREOF

FILE REFERENCE: CL000929

CURRENT APPLICATION NUMBER: US/60/243,468

CURRENT FILING DATE: 2000-10-27

NUMBER OF SEQ ID NOS: 2121

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1271

LENGTH: 685

TYPE: PRT

ORGANISM: HUMAN

US-60-243-468-1271

Query Match 88.8%; Score 1575; DB 27; Length 685;
Best Local Similarity 90.0%; Pred. No. 1.5e-122;
Matches 323; Conservative 0; Mismatches 0; Indels 36; Gaps 1;

QY 1 GTLVTKVAPVSAPPKVSSGPRLPAPQIVAVKAPNTTTIQFPANLQLPPGTVLIKSNSGPL 60
Db 67 GTLVTKVAPVSAPPKVSSGPRLPAPQIVAVKAPNTTTIQFPANLQLPPGTVLIKSNSGPL 126
QY 61 MLVSPQQTVTTRAETTSNITSRPAPVNPQTVKICTVPSNSQLIKKVAVTPVKKLAQIGT 120
Db 127 MLVSPQQTVTTRAETTSNITSRPAPVNPQTVKICTVPSNSQLIKKVAVTPVKKLAQIGT 163
QY 121 TVVTTVPKPSVQSVAVPTSVVTVTPGKPLNTVTTLPKPSLSGASSTPSNEPNAENSA 180
Db 164 -----SVAVPTSVVTVTPGKPLNTVTTLPKPSLSGASSTPSNEPNAENSA 210

QY 181 VOINLSPMTLENNKCNFLAMLIKACSGSPENGVNKKLVBLQDLDAKIEAEFFTRK 240
DB 211 VOINLSPMTLENNKCNFLAMLIKACSGSPENGVNKKLVBLQDLDAKIEAEFFTRK 270
QY 241 LVYELKSSPQHLVFLPKKSVVALRQLLPNSQSFQOCVQOQTSSDMVIATCTTTTSPV 300
DB 271 LVYELKSSPQHLVFLPKKSVVALRQLLPNSQSFQOCVQOQTSSDMVIATCTTTTSPV 330
QY 301 VTTTSSSSQSEKSIIVSGATAPRTVSOTLNLAPGVGAKAGVWTLHSGVGPAAATGGTT 359
DB 331 VTTTSSSSQSEKSIIVSGATAPRTVSOTLNLAPGVGAKAGVWTLHSGVGPAAATGGTT 389

RESULT 6
PCT-US94-01114-16
; Sequence 16, Application PC/TUS9401114
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAPS AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOUBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01114
; FILING DATE: 28-JAN-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: FP-57650-2/AUT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 737 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-01114-16

Query Match 27.7%; Score 491; DB 1; Length 737;
Best Local Similarity 38.4%; Pred. No. 1.4e-31;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;
QY 13 PKKVSSG-----PRLPAPQIVAVKAPNTTITQFPANLQLPFGTVLKSSGGLM 61
DB 58 PTATSGIRATLPTVLAPRLPQ-----PONPTNIQ---NFQPPGMVLRSENGQLL 118
QY 62 LVSPQQTVTR-----AETTSNITSRPVAVPANTQVTKICTVPSNSSOLIKKVAVTPVKKL 115
DB 119 MI-PQALAQMOAQAHAQPTTMAPRPATPTSPAPVQISTVQAPCTPIAR-QVTP---- 172

QY 116 AOIGTTVTVTVPKPSVQSVAVPTSVVTVPGKPLNTVT--TLKPSSLGASS-----TPS 168
DB 173 ----TTIIKQV---SQAQTTVQPSATLQSRPGVQPOLVLGGAAQTASLGTATAVQTGTQ 225
QY 169 NE-PNLKAENSAVQINLSPTMLENNKCNFLAMLIKACSGSPENGVNKKLVBLQDL 227
DB 226 RTVPGATTTSSAATE-----TMENYKCKNFLLSTLIKASSGKQSTETAANYKELVQNL 279
QY 228 LDKATEAEFFTRKLVYELKSSPQHLVFLPKKSVVALRQLLPNSQSFQOCVQO-----TS 283
DB 280 LDKTEAEDEFTSRLRELNSPQPLVFLPKKSLPALROLTLPDSAAFIQSSQOQPPPTPS 339
QY 284 SDMWIATCTTTTSPVTTTSSQSEKSIIVSGATAPRTVSV 327
DB 340 Q-----ATTALTAVLSSSVQRTAGKTAATVTSALQPPVLSL 376

RESULT 7
PCT-US01-20592-14
; Sequence 14, Application PC/TUS0120592
; GENERAL INFORMATION:
; APPLICANT: ANADYS PHARMACEUTICALS, INC.
; APPLICANT: THOMPSON, Craig
; APPLICANT: MOORE, Jeffrey
; APPLICANT: BURMAN, Ed T.
; APPLICANT: BRADLEY, John
; APPLICANT: DESILVA, Thamara
; APPLICANT: HARRIS, Sandra
; APPLICANT: KOMARNITSKY, Svetlana
; APPLICANT: MENDILLO, Marc
; APPLICANT: MOORE, Daniel
; APPLICANT: MCCOY, Melissa
; APPLICANT: SANDERSON, Karen
; APPLICANT: HAQ, Tariq
; APPLICANT: ZHU, Shuhao
; APPLICANT: LONG, Fan
; APPLICANT: DAVIDOV, Eugene
; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
; FILE REFERENCE: 0342/1G548-US1
; CURRENT APPLICATION NUMBER: PCT/US01/20592
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,164
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/224,457
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 1023
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: human genbank accession #: CAA72189
; NAME/KEY: misc.feature
; OTHER INFORMATION: Corresponds to SEQ ID NO: 87
PCT-US01-20592-14

Query Match 27.7%; Score 491; DB 1; Length 1023;
Best Local Similarity 38.4%; Pred. No. 2.3e-31;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;
QY 13 PKKVSSG-----PRLPAPQIVAVKAPNTTITQFPANLQLPFGTVLKSSGGLM 61
DB 414 PTATSGIRATLPTVLAPRLPQ-----PONPTNIQ---NFQPPGMVLRSENGQLL 464
QY 62 LVSPQQTVTR-----AETTSNITSRPVAVPANTQVTKICTVPSNSSOLIKKVAVTPVKKL 115
DB 465 MI-PQALAQMOAQAHAQPTTMAPRPATPTSPAPVQISTVQAPCTPIAR-QVTP---- 518
QY 116 AOIGTTVTVTVPKPSVQSVAVPTSVVTVPGKPLNTVT--TLKPSSLGASS-----TPS 168
DB 519 ----TTIIKQV---SQAQTTVQPSATLQSRPGVQPOLVLGGAAQTASLGTATAVQTGTQ 571

QY 169 NE-PNLKAENSAVAQINLSPMTLENVKKCNFLAMLIKACSGSQSPMGONVKKLYEQL 227
DB 572 RTVPGATTSSAAE-----TMENVKKCKNFTSLTIKLASSGKQSTETAAANVKELVQNL 625
QY 228 LDAKIEAEETRKLYVELKSSQPOHLYPFLKKSVALRQLLPNSQSFIOQCQVQ-----TS 283
DB 626 LDGKIEAEDFTSRILYRELNSSPQPLVPFLKRSIPALRQLTPDPSAAFIQSQOQPPPTTS 685
QY 284 SDMYIATCTTTVTSPVVTITVSSQSEKSIIVSGATAPRTVS 327
DB 686 Q-----ATTALTAVLSSSVQRTAGTAAATVTSALQPPVL 722

RESULT 8

PCT-US01-20592A-14

; Sequence 14, Application PC/TUS0120592A

; GENERAL INFORMATION:

; APPLICANT: ANADYS PHARMACEUTICALS, INC.

; APPLICANT: THOMPSON, Craig

; APPLICANT: MOORE, Jeffrey

; APPLICANT: BUURMAN, Ed T.

; APPLICANT: BRADLEY, John

; APPLICANT: DESILVA, Thamara

; APPLICANT: HARRIS, Sandra

; APPLICANT: KOMARNITSKY, Svetlana

; APPLICANT: MENDILLO, Marc

; APPLICANT: MOORE, Daniel

; APPLICANT: MCCOY, Melissa

; APPLICANT: SANDERSON, Karen

; APPLICANT: HAQ, Tariq

; APPLICANT: ZHU, Shuhao

; APPLICANT: LONG, Fan

; APPLICANT: DAVIDOV, Eugene

; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE

; FILE REFERENCE: 0342/1G548-US1

; CURRENT APPLICATION NUMBER: PCT/US01/20592A

; CURRENT FILING DATE: 2001-06-28

; PRIOR APPLICATION NUMBER: US 60/215,164

; PRIOR FILING DATE: 2000-06-29

; PRIOR FILING DATE: 2000-08-10

; NUMBER OF SEQ ID NOS: 146

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 14

; LENGTH: 1023

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; OTHER INFORMATION: Corresponds to SEQ ID NO: 87

; PCT-US01-20592A-14

Query Match 27.7%; Score 491; DB 1; Length 1023;
Best Local Similarity 38.4%; Pred. No. 2.3e-31;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

QY 13 PKVSSG-----PRLPAPQIVAVKAPNTTTTOFPANLQLPPTGVLIKNSGGLM 61
DB 414 PTATTSIGIRATLTPTVLAPRLPQ-----PQNPNTIQ---NFQLPPGMVLVRSNGQL 464
QY 62 LVSPQQTVTR-----AETTSNITSRAVPANPOTVKICITVPNSSQLIKKAVTPVKKL 115
DB 465 MI-PQALAQMOAQAHAQPTMAPRATPTSAPPVQISTVQAPGTPIIAR-QVTP----- 518
QY 116 AQIGTTVVTTVPKSSVQSVAVPTSVVTPGKPLNTVT---TLKPSSILGASS-----TPS 168
DB 519 -----TTIIKQV---SQAQTTVPQSATLQSPGVQPOLVLCGAAGTASLCTATATVQTGP 571
QY 169 NE-PNLKAENSAVAQINLSPMTLENVKKCNFLAMLIKACSGSQSPMGONVKKLYEQL 227
DB 572 RTVPGATTSSAAE-----TMENVKKCKNFTSLTIKLASSGKQSTETAAANVKELVQNL 625

QY 228 LDAKIEAEETRKLYVELKSSQPOHLYPFLKKSVALRQLLPNSQSFIOQCQVQ-----TS 283
DB 626 LDGKIEAEDFTSRILYRELNSSPQPLVPFLKRSIPALRQLTPDPSAAFIQSQOQPPPTTS 685
QY 284 SDMYIATCTTTVTSPVVTITVSSQSEKSIIVSGATAPRTVS 327
DB 686 Q-----ATTALTAVLSSSVQRTAGTAAATVTSALQPPVL 722

RESULT 9

US-09-893-519A-14

; Sequence 14, Application US/09893519A

; GENERAL INFORMATION:

; APPLICANT: ANADYS PHARMACEUTICALS, INC.

; APPLICANT: THOMPSON, Craig

; APPLICANT: MOORE, Jeffrey

; APPLICANT: BUURMAN, Ed T.

; APPLICANT: BRADLEY, John

; APPLICANT: DESILVA, Thamara

; APPLICANT: HARRIS, Sandra

; APPLICANT: KOMARNITSKY, Svetlana

; APPLICANT: MENDILLO, Marc

; APPLICANT: MOORE, Daniel

; APPLICANT: MCCOY, Melissa

; APPLICANT: SANDERSON, Karen

; APPLICANT: HAQ, Tariq

; APPLICANT: ZHU, Shuhao

; APPLICANT: LONG, Fan

; APPLICANT: DAVIDOV, Eugene

; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE

; FILE REFERENCE: 0342/1G548-US2

; CURRENT APPLICATION NUMBER: US/09/893,519A

; CURRENT FILING DATE: 2001-06-28

; PRIOR APPLICATION NUMBER: US 60/215,164

; PRIOR FILING DATE: 2000-06-29

; PRIOR FILING DATE: 2000-08-10

; NUMBER OF SEQ ID NOS: 146

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 14

; LENGTH: 1023

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; OTHER INFORMATION: Corresponds to SEQ ID NO: 87

; PUBLICATION INFORMATION:

; DATABASE ACCESSION NUMBER: Human Genbank/CAA72189

; DATABASE ENTRY DATE: 1997-06-25

; RELEVANT RESIDUES: (1)..(1023)

; US-09-893-519A-14

Query Match 27.7%; Score 491; DB 22; Length 1023;
Best Local Similarity 38.4%; Pred. No. 2.3e-31;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

QY 13 PKVSSG-----PRLPAPQIVAVKAPNTTTTOFPANLQLPPTGVLIKNSGGLM 61
DB 414 PTATTSIGIRATLTPTVLAPRLPQ-----PQNPNTIQ---NFQLPPGMVLVRSNGQL 464
QY 62 LVSPQQTVTR-----AETTSNITSRAVPANPOTVKICITVPNSSQLIKKAVTPVKKL 115
DB 465 MI-PQALAQMOAQAHAQPTMAPRATPTSAPPVQISTVQAPGTPIIAR-QVTP----- 518
QY 116 AQIGTTVVTTVPKSSVQSVAVPTSVVTPGKPLNTVT---TLKPSSILGASS-----TPS 168
DB 519 -----TTIIKQV---SQAQTTVPQSATLQSPGVQPOLVLCGAAGTASLCTATATVQTGP 571
QY 169 NE-PNLKAENSAVAQINLSPMTLENVKKCNFLAMLIKACSGSQSPMGONVKKLYEQL 227
DB 572 RTVPGATTSSAAE-----TMENVKKCKNFTSLTIKLASSGKQSTETAAANVKELVQNL 625
QY 228 LDAKIEAEETRKLYVELKSSQPOHLYPFLKKSVALRQLLPNSQSFIOQCQVQ-----TS 283

```

Db 626 LDKIEAEFTSLRYRELNSSQPQYLVFLKRLPALRLQTLTPDSAAFIQSQSQPPPTPS 685
QY 284 SDMWIATCTTIVTSPVVTSSSSQSEKSIIVSGATAPRTVSV 327
Db 686 Q-----ATTALTAVLSSSVQRTAGKTAAVTTSALQPPVL 722

RESULT 10
PCT-US02-30474-1660
; Sequence 1660, Application PC/TUS0230474
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Ghosh, Malabika
; APPLICANT: Wang, Dunrui
; APPLICANT: Ma, Yunqing
; APPLICANT: Asundi, Vinod
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wang, Gezhi
; APPLICANT: Halsey-Vicente, Dana
; APPLICANT: Drmanac, Radoje T
; TITLE OF INVENTION: Novel Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 810CIP PCT
; CURRENT APPLICATION NUMBER: PCT/US02/30474
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 60/324,631
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 3476
; SOFTWARE: pt_FL_genes Version 6.0
; SEQ ID NO 1660
; LENGTH: 1051
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-30474-1660

Query Match 27.7%; Score 491; DB 1; Length 1051;
Best Local Similarity 38.4%; Pred. No. 2.4e-31;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

QY 13 PKVSSG-----PRLPAPQIVAVKAPNTTTTIOFPANLQPPGTVLKNSGPI 61
Db 382 PRATTSIGIRATLTPVLAPRLPQ-----PQNTNIQ---NFQLPPGMVLVRSENGQL 432
QY 62 LVSPQQTTR-----AETTSNITSRPVAPNPQVKTCTVPSNSSOLIKKVVAVTPVK 115
Db 433 MI-PQALAQMAQAHAQAPQTTMAPRPPTSPAPVQISTVQAPGTPIAR-QVTP---- 486
```

```

QY 116 AQTGTTVTVTKPSSVQSVAVPTSIVTTPGKPLNTVT--TLKPSLSGASS-----TPS 168
Db 487 ----TTIIKQV---SOAQTVPQSPATLQSRPGVQPOLVLGGAQAATSLGTATAVQTGPQ 539
QY 169 NE-PNLKANSAAVQINLSPTMLENVKCKNFAMLIKACSGSQSPKMGQNVKYLVEQL 227
Db 540 RTVPGATTTSSAATE-----TMENVKCKNFLESTLIKASSGKQSTETAANYKELVQNL 593
QY 228 LDKIEAEFTSLRYRELNSSQPQYLVFLKRLPALRLQTLTPDSAAFIQSQSQPPPTPS 283
Db 594 LDKIEAEFTSLRYRELNSSQPQYLVFLKRLPALRLQTLTPDSAAFIQSQSQPPPTPS 653
QY 284 SDMWIATCTTIVTSPVVTSSSSQSEKSIIVSGATAPRTVSV 327
Db 654 Q-----ATTALTAVLSSSVQRTAGKTAAVTTSALQPPVL 690

RESULT 11
US-60-324-631-1665
; Sequence 1665, Application US/60324631
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Ghosh, Malabika
; APPLICANT: Wang, Dunrui
; APPLICANT: Ma, Yunqing
; APPLICANT: Asundi, Vinod
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wang, Gezhi
; APPLICANT: Halsey-Vicente, Dana
; APPLICANT: Drmanac, Radoje T
; TITLE OF INVENTION: Novel Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 810
; CURRENT APPLICATION NUMBER: US/60/324,631
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: US 09/577,409
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: PCT/US01/04927
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 09/519,705
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/574,454
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: PCT/US01/04941
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 09/540,217
; PRIOR FILING DATE: 2000-03-31
```

FILE REFERENCE: 810CIP PCT
CURRENT APPLICATION NUMBER: PCT/US02/30474
CURRENT FILING DATE: 2002-09-24
PRIOR APPLICATION NUMBER: US 60/324,631
PRIOR FILING DATE: 2001-09-24
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02623
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: US 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: PCT/US01/03800
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 09/515,126
PRIOR FILING DATE: 2000-02-28
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 3476
SOFTWARE: pc_flg_genes Version 6.0
SEQ ID NO 3271
LENGTH: 1083
TYPE: PRT
ORGANISM: Homo sapiens
NAME/KEY: misc_feature
LOCATION: (1)...(1083)
OTHER INFORMATION: Xaa = any amino acid or symbol as shown in the table 8 as set
PCT-US02-30474-3271

Query Match 27.7%; Score 491; DB 27; Length 1051;
Best Local Similarity 38.4%; Pred. No. 2.4e-31;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

Y 13 PKVSSG-----PRLPAPQIVAVKAPNTTIOFPANLQLPCTGVLIKNSGGLM 61
D 382 PTATTSIGIRATLTPTVLAPRLPQ-----PQNPTNIO---NFQLPPGMVLVRSNGQLL 432
Y 62 LVSPQQTVTR-----AETTSNISRPAVPANPOTVKICTVPSNSSOLIKKAVTPVKKL 115
D 433 MI-FQALAAQQAHAQOQPMAPRPATPSAPVQISTVQAGTPIIAR-QVTP-----486
Y 116 AQIGTVVTVTP-----KPSVQSVAVPTSVVTP-----TLKPSSLGASS-----TPS 168
D 487 ---TTIKQV---SQAGTVQPSATLQSPGVQPOLVGLGAAQTASLGATATVATGTPQ 539
Y 169 NE-PNLKAENSAAVQINLSPTMLENVKCKNFAMLIKACSGSQSPENGVKLVQL 227
D 540 RTVPGATTSSAATE-----TMENVKCKNFSLTLIKLASGSGQSTETAANVKELVQNL 593
Y 228 LDKIAEAEFTFRKLYVELKSSQPHLVPFLKSVVALRQLLPNSQSFIOQCVOQ-----TS 283
D 594 LDGKIEADFTSRLYRELNSPQYLVPLFKRSPLALRQLTPDPSAAFIQSQSQQPPPTTS 653
Y 284 SDMWIATCTTTTTSPPVTVTVSSSQSEKSIIVSGATAPRTVS 327
D 654 Q-----ATTALTAVLSSSVQRTAGTAAVTATVTSALQPPVL 690

RESULT 12
PCT-US02-30474-3271
Sequence 3271, Application PC/TUS0230474
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Xue, Aigong J.
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Jian-Rui
APPLICANT: Wehrman, Tom
APPLICANT: Zhou, Ping
APPLICANT: Ghosh, Malabika
APPLICANT: Wang, Dunrui
APPLICANT: Ma, Yunqing
APPLICANT: Asundi, Vinod
APPLICANT: Wang, Zhiwei
APPLICANT: Meng, Gezhi
APPLICANT: Haley-Vicente, Dana
APPLICANT: Drmanac, Radoje T
TITLE OF INVENTION: Novel Nucleic Acids and
TITLE OF INVENTION: Polypeptides

Query Match 27.0%; Score 479; DB 1; Length 1083;
Best Local Similarity 36.3%; Pred. No. 2.6e-30;

Matches 12; Conservative 50; Mismatches 97; Indels 76; Gaps 12;

Y 13 PKVSSG-----PRLPAPQIVAVKAPNTTIOFPANLQLPCTGVLIKNSGGLM 61
D 414 PTATTSIGIRATLTPTVLAPRLPQ-----PQNPTNIO---NFQLPPGMVLVRSNGQLL 464
Y 62 LVSPQQTVTR-----AETTSNISRPAVPANPOTVKICTVPSNSSOLIKKAVTPVKKL 115
D 465 MI-FQALAAQQAHAQOQPMAPRPATPSAPVQISTVQAGTPIIAR-QVTP-----522
Y 116 AQIGTVVTVTP-----KPSVQSVAVPTSVVTP-----GKPLNTVTVTLKPSL 161
D 523 SRVSOAGTVQPSATLKRSPGQPOLVGLGAAQTASLGATATVATGTPQRTVTP-----575
Y 162 GASSTPSNEPNLKAENSAAVQINLSPTMLENVKCKNFAMLIKACSGSQSPENGVK 221
D 576 GATTSS-----VARETMENVKCKNFSLTLIKLASGSGQSTETAANVK 619
Y 222 KLVQLDLAKIAEAEFTFRKLYVELKSSQPHLVPFLKSVVALRQLLPNSQSFIOQCVOQ 281
D 620 ELVONLDGKIEADFTSRLYRELNSPQYLVPLFKRSPLALRQLTPDPSAAFIQSQSQQ 679
Y 282 ---TSSDMVIATCTTTTTSPPVTVTVSSSQSEKSIIVSGATAPRTVS 327
D 680 PPPPTSQ-----ATTALTAVLSSSVQRTAGTAAVTATVTSALQPPVL 722

RESULT 13

US-60-167-217-11401
Sequence 11401, Application US/60167217
GENERAL INFORMATION:
APPLICANT: Li, Peter W. D.
TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
TITLE OF INVENTION: THEREOF

FILE REFERENCE: CL000152
CURRENT APPLICATION NUMBER: US/60/167,217
CURRENT FILING DATE: 1999-11-24
NUMBER OF SEQ ID NOS: 23195
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11401
LENGTH: 899
TYPE: PRT
ORGANISM: Drosophila
US-60-167-217-11401

Query Match 12.3%; Score 218; DB 27; Length 899;
Best Local Similarity 24.9%; Pred. No. 1.6e-08;
Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;

QY 25 POIVAVKAPNTTIOFPANLQPPGTVLKNSGPMML-----LVKTDNGFQLLRVGTGTTGPTTVTQITNTSNN 162
DB 111 POSPSITLSTLTGQTPA-----LVKTDNGFQLLRVGTGTTGPTTVTQITNTSNN 162
QY 77 NITSRPAVPANPQTVKICIVPNSSSQ-----LIKKVAVTPVKKLAQIGTIVVTVTP 127
DB 163 NITSTNHNHTTQ-IRLQVPAASMTNTATSNIIIVSVASSGYANSQPPHLLTQNAQ 221
QY 128 KPSSVQSVAVPTSVVTVTPGKPLNTVTLKPSLSGASSTPSNEPNLKAENSAVQINLSP 187
DB 222 APQLPQITQIOTIPAQSOQQOQVNVSSAGGTATAVSSTA-----ATT 265
QY 188 TMLENVK-KCKNFLAMLIKACSGSPGEMGVKLVQQLDADAKTEAEFTRKLYVELK 246
DB 266 TOGNTKECKRFLANLIEL--STREPKEVKNVRLIOELVNAVPEPEFCDRLERLN 323
QY 247 SSPQHLVFPFKKSVVALRQL-----LPNSQSFTQ--- 276
DB 324 ASPQCLIGFLKSLPLLRQALYKELVIEGKPPQHVGLAGLSQQLPKIQAIQIRPIG 383
QY 277 ----QCVOQTSSDMVIACTTTTSPVTVTVSSSQSEKSIIVSGATAPRTVS---VQT 329
DB 384 PSQTTTIGTQVRMI--TPNALGTTPRTTIGHTTISKQPPN---IRLPTAPRLVNTGGIRT 438
QY 330 LNPLAGPVGAKAGVTVLHSGVPTAA 354
DB 439 QIP-SLOVPGQANIVQIR--GPQHA 460

RESULT 14

US-60-173-464-9223
Sequence 9223, Application US/60173464

GENERAL INFORMATION:
APPLICANT: LI, Peter W.D.
TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
FILE REFERENCE: CL000173
CURRENT APPLICATION NUMBER: US/60/173,464
CURRENT FILING DATE: 1999-12-29
NUMBER OF SEQ ID NOS: 30269
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9223
LENGTH: 921
TYPE: PRT
ORGANISM: Drosophila
US-60-173-464-9223

Query Match 12.3%; Score 218; DB 27; Length 921;
Best Local Similarity 24.9%; Pred. No. 1.6e-08;
Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;

QY 25 POIVAVKAPNTTIOFPANLQPPGTVLKNSGPMML-----LVKTDNGFQLLRVGTGTTGPTTVTQITNTSNN 184
DB 133 POSPSITLSTLTGQTPA-----LVKTDNGFQLLRVGTGTTGPTTVTQITNTSNN 184
QY 77 NITSRPAVPANPQTVKICIVPNSSSQ-----LIKKVAVTPVKKLAQIGTIVVTVTP 127

DB 185 NITSTNHNHTTQ-IRLQVPAASMTNTATSNIIIVSVASSGYANSQPPHLLTQNAQ 243
QY 128 KPSSVQSVAVPTSVVTVTPGKPLNTVTLKPSLSGASSTPSNEPNLKAENSAVQINLSP 187
DB 244 APQLPQITQIOTIPAQSOQQOQVNVSSAGGTATAVSSTA-----ATT 287
QY 188 TMLENVK-KCKNFLAMLIKACSGSPGEMGVKLVQQLDADAKTEAEFTRKLYVELK 246
DB 288 TOGNTKECKRFLANLIEL--STREPKEVKNVRLIOELVNAVPEPEFCDRLERLN 345
QY 247 SSPQHLVFPFKKSVVALRQL-----LPNSQSFTQ--- 276
DB 346 ASPQCLIGFLKSLPLLRQALYKELVIEGKPPQHVGLAGLSQQLPKIQAIQIRPIG 405
QY 277 ----QCVOQTSSDMVIACTTTTSPVTVTVSSSQSEKSIIVSGATAPRTVS---VQT 329
DB 406 PSQTTTIGTQVRMI--TPNALGTTPRTTIGHTTISKQPPN---IRLPTAPRLVNTGGIRT 460
QY 330 LNPLAGPVGAKAGVTVLHSGVPTAA 354
DB 461 QIP-SLOVPGQANIVQIR--GPQHA 482

RESULT 15

US-60-173-464-20611
Sequence 20611, Application US/60173464

GENERAL INFORMATION:
APPLICANT: LI, Peter W.D.
TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
FILE REFERENCE: CL000173
CURRENT APPLICATION NUMBER: US/60/173,464
CURRENT FILING DATE: 1999-12-29
NUMBER OF SEQ ID NOS: 30269
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 20611
LENGTH: 921
TYPE: PRT
ORGANISM: Drosophila
US-60-173-464-20611

Query Match 12.3%; Score 218; DB 27; Length 921;
Best Local Similarity 24.9%; Pred. No. 1.6e-08;
Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;

QY 25 POIVAVKAPNTTIOFPANLQPPGTVLKNSGPMML-----LVKTDNGFQLLRVGTGTTGPTTVTQITNTSNN 184
DB 133 POSPSITLSTLTGQTPA-----LVKTDNGFQLLRVGTGTTGPTTVTQITNTSNN 184
QY 77 NITSRPAVPANPQTVKICIVPNSSSQ-----LIKKVAVTPVKKLAQIGTIVVTVTP 127
DB 185 NITSTNHNHTTQ-IRLQVPAASMTNTATSNIIIVSVASSGYANSQPPHLLTQNAQ 243
QY 128 KPSSVQSVAVPTSVVTVTPGKPLNTVTLKPSLSGASSTPSNEPNLKAENSAVQINLSP 187
DB 244 APQLPQITQIOTIPAQSOQQOQVNVSSAGGTATAVSSTA-----ATT 287
QY 188 TMLENVK-KCKNFLAMLIKACSGSPGEMGVKLVQQLDADAKTEAEFTRKLYVELK 246
DB 288 TOGNTKECKRFLANLIEL--STREPKEVKNVRLIOELVNAVPEPEFCDRLERLN 345
QY 247 SSPQHLVFPFKKSVVALRQL-----LPNSQSFTQ--- 276
DB 346 ASPQCLIGFLKSLPLLRQALYKELVIEGKPPQHVGLAGLSQQLPKIQAIQIRPIG 405
QY 277 ----QCVOQTSSDMVIACTTTTSPVTVTVSSSQSEKSIIVSGATAPRTVS---VQT 329
DB 406 PSQTTTIGTQVRMI--TPNALGTTPRTTIGHTTISKQPPN---IRLPTAPRLVNTGGIRT 460
QY 330 LNPLAGPVGAKAGVTVLHSGVPTAA 354
DB 461 QIP-SLOVPGQANIVQIR--GPQHA 482

Search completed: February 16, 2003, 22:18:31
Job time : 133.688 secs

This Page Blank (uspto)

; Sequence 63929, Application US/09724676A

```
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 63929
; LENGTH: 679
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-63929

Query Match      27.7%; Score 491; DB 5; Length 679;
Best Local Similarity 38.4%; Pred. No. 4,8e-21;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14

Y   13 PEKVSSG-----PRLPAQIVAAKAPNTTIOFPANQLPPGVILKNSGPILM 61
       |::||| ||||| |::| |::| |::| |::| |::| |::| |::| |::|
Db  286 PLATISGRATLTPVYLARLPQ-----PONPTIQ--NFQLPGNAVLRSENGQL 336
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Y   62 LVSPQQTIVTR-----AETTSNITSRPAYPANPQTVKICTVPNSSQLIKKYAVTPVKKL 115
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  337 MI-PQALAQMOQAHAQOQTMMARPAPTASAPVQISTVOAGCPRIIAR-QVTP----- 390
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Y   116 AOGITVTYTVKKPSSVGSAVPTSIVTYTPCKPLNTV--TLKPSSLCASS-----TPS 168
       391 ----TTIRKQV---SOAOITVPSATILRSGPOPOLVLGAAGVRSIGTAATAVGTGPQ 443
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Y   169 NE-PNLKAENSAVAOVQINSPTMLLENVKKCKKNFLAMLIKLAGSGOSPEKGONVKRYEOL 227
       |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  444 RVPAPATTSSAATE-----TMENYKCKKNLTSLIKLASGKOSTETAANKVELVOLN 497
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Y   228 LDKATEAEERTKRLLYELEKSSPQPHLYPELPKRSVALROLLEPNSSOFIOCVQO----TS 283
       ||||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  498 LDGKTEADEFTSRILELNSSPQYLVPLPKLSPLRALROLTPDSAAFIOQSQQQPPPTSS 557
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Y   284 SDMYATCTTYYTSPVYTITVSSSGSEKSIIYSGATPRATYSV 327
       ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  558 Q-----ATTALTAVLVSSSVORTAGTKATYATSTALOPLVLT 594
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 3
US-09-724-676-63927
; Sequence 63927, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 63927
; LENGTH: 709
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-63927

Query Match      27.7%; Score 491; DB 5; Length 709;
Best Local Similarity 38.4%; Pred. No. 5e-21;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14

Y   13 PEKVSSG-----PRLPAQIVAAKAPNTTIOFPANQLPPGVILKNSGPILM 61
       |::||| ||||| |::| |::| |::| |::| |::| |::| |::| |::|
Db  286 PLATISGRATLTPVYLARLPQ-----PONPTIQ--NFQLPGNAVLRSENGQL 336
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Y   62 LVSPQQTIVTR-----AETTSNITSRPAYPANPQTVKICTVPNSSQLIKKYAVTPVKKL 115
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  337 MI-PQALAQMOQAHAQOQTMMARPAPTASAPVQISTVOAGCPRIIAR-QVTP----- 390
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Y   116 AOGITVTYTVKKPSSVGSAVPTSIVTYTPCKPLNTV--TLKPSSLCASS-----TPS 168
       391 ----TTIRKQV---SOAOITVPSATILRSGPOPOLVLGAAGVRSIGTAATAVGTGPQ 443
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Y   169 NE-PNLKAENSAVAOVQINSPTMLLENVKKCKKNFLAMLIKLAGSGOSPEKGONVKRYEOL 227
       |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  444 RVPAPATTSSAATE-----TMENYKCKKNLTSLIKLASGKOSTETAANKVELVOLN 497
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Y   228 LDKATEAEERTKRLLYELEKSSPQPHLYPELPKRSVALROLLEPNSSOFIOCVQO----TS 283
       ||||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  498 LDGKTEADEFTSRILELNSSPQYLVPLPKLSPLRALROLTPDSAAFIOQSQQQPPPTSS 557
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Y   284 SDMYATCTTYYTSPVYTITVSSSGSEKSIIYSGATPRATYSV 327
       ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  558 Q-----ATTALTAVLVSSSVORTAGTKATYATSTALOPLVLT 594
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

```

Db      391  ----TTIIKQV---SQAQTTVQPSAATLQNSPCGVQPOLVLGGAAGTASLSLTATTAATVQCTGTQ  443
OY      169  NE-PLUKAENSAVOINLSPTMLENVKKCKKNFLAMLIKIACSGSOSPEMGQNVKLIQLEQ  227
Db      444  RTVPGATTTSSAATE-----TMENVKCKCKNFLTSLIKIACSGKOSTETAAVANKELVQNL  497
OY      228  LDKAEAEFFTKRLVLELKSSPQPHLYPPLKKSVAALRDLRNSQSFIOOCVQO-----TS  283
Db      498  LDGKIEAEDEFTSLRELKNSSPQPLVPLPKRSPLALRDLTPDSAAFIOOQSOQPPPPPTS  557
OY      284  SDMWATCTCTVTTSPVYTTVTTSSQSEKSIIVSGATAPRTVS  327
Db      558  Q-----ATTALTAIVLSSSVQRTAGKTAATVTSALQPPVLSL  594

RESULT 4
US-09-724-676A-63927
: Sequence 63927, Application US/09724676A
: GENERAL INFORMATION:
: APPLICANT: Comugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Comugen
: CURRENT APPLICATION NUMBER: US/09/724,676A
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 63927
: LENGTH: 709
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-724-676A-63927

Query Match      27.7%; Score 491; DB 5; Length 709;
Best Local Similarity 38.4%; Pred. No. 5e-21;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14

OY      13  PKVSSG-----PRLEAPQIVAKAANTTTTIOPPANLQLPFGVILIKNSGLM  61
Db      286  PTATTSIGRATLTPVLAPRLPQ-----PONPTNIQ--NFOJPPGMVLVRESENGQL  336
OY      62  LVSPQOTVTR-----AETTSNITSBPAPAPAPQVYKICTVNSSQLIKKVAATPVYK  115
Db      337  MI-PQALAAQOAOAHQOQTMMAPRPATVTSAPPQISTVOAPGPIIAR-QVTP----  390
OY      116  AQIGTTVYTVPKPSSVQSAVPTSVYTTPGKPLNTVY--TLKPSLSGASS-----TPS  168
Db      391  ----TTIIKQV---SQAQTTVQPSAATLQNSPCGVQPOLVLGGAAGTASLSLTATTAATVQCTGTQ  443
OY      169  NE-PLUKAENSAVOINLSPTMLENVKKCKKNFLAMLIKIACSGSOSPEMGQNVKLIQLEQ  227
Db      444  RTVPGATTTSSAATE-----TMENVKCKCKNFLTSLIKIACSGKOSTETAAVANKELVQNL  497
OY      228  LDKAEAEFFTKRLVLELKSSPQPHLYPPLKKSVAALRDLRNSQSFIOOCVQO-----TS  283
Db      498  LDGKIEAEDEFTSLRELKNSSPQPLVPLPKRSPLALRDLTPDSAAFIOOQSOQPPPPPTS  557
OY      284  SDMWATCTCTVTTSPVYTTVTTSSQSEKSIIVSGATAPRTVS  327
Db      558  Q-----ATTALTAIVLSSSVQRTAGKTAATVTSALQPPVLSL  594

RESULT 5
US-09-724-676-63920
: Sequence 63920, Application US/09724676
: GENERAL INFORMATION:
: APPLICANT: Comugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Comugen
: CURRENT APPLICATION NUMBER: US/09/724,676
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2

```



```

Db 433 MI-PQALAQMOAHAPQTTMAPRATPTTSAPPVQISTVQAPCTPIIAR-QVTP----- 486
Qy 116 AOGITVTVTPKPSVSVAVPTSVTVTPGKPLNTVT--TLKPSIGASS-----TPS 168
Db 487 -----TTIIKQV-----SQAQTTVQPSATLQSPGVQPOLVLGGAQATASIGTATAVQGTGPQ 539
Qy 169 NE-PLKAKENSAVOINLSPTMLENVKCKNFMLILKLACSGSGSPMGONVKKLVEQL 227
Db 540 RTVPGATTSSAATE-----TMENVKCKNFSLTLILKLACSGSGKOSTETAAVNEKLVONL 593
Qy 228 LDKATEAEFTFRKLYVELKSSPOPILVPELKKSVVALROLPLNSQSFIOQCVOQ-----TS 283
Db 594 LDGKTEADEFTSRLYRELNSSPOPILVPELKRSLPALROLPLPDSNAFTQSQOQPPPTTS 653
Qy 284 SDMWIATCTTVTTSPTVTVSSQSEKSIIVSGATAPRTVS 327
Db 654 Q-----ATTALTAIVLVSSVQRTAGKTAATVTSALOPPVLSL 690

RESULT 12
US-09-724-676A-63921
; Sequence 63921, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63921
; LENGTH: 805
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-63921

Query Match
Best Local Similarity 38.4%; Score 491; DB 5; Length 805;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

Qy 13 PPKVSSG-----PRLPAPQIVAVKAPNTTTFPANLQLPCTVLIKSNSGPLM 61
Db 382 PTATTSGRATLTPVLAPRLPQ-----PONPTNIO-----NFQLPQGMVLVRSENGQL 432
Qy 62 LVSPQQTIVTR-----AETTSNITSRPVAVPANPQTVKICITVPNSSQILKKVAVTPVKKL 115
Db 433 MI-PQALAQMOAHAPQTTMAPRATPTTSAPPVQISTVQAPCTPIIAR-QVTP----- 486
Qy 116 AOGITVTVTPKPSVSVAVPTSVTVTPGKPLNTVT--TLKPSIGASS-----TPS 168
Db 487 -----TTIIKQV-----SQAQTTVQPSATLQSPGVQPOLVLGGAQATASIGTATAVQGTGPQ 539
Qy 169 NE-PLKAKENSAVOINLSPTMLENVKCKNFMLILKLACSGSGSPMGONVKKLVEQL 227
Db 540 RTVPGATTSSAATE-----TMENVKCKNFSLTLILKLACSGSGKOSTETAAVNEKLVONL 593
Qy 228 LDKATEAEFTFRKLYVELKSSPOPILVPELKKSVVALROLPLNSQSFIOQCVOQ-----TS 283
Db 594 LDGKTEADEFTSRLYRELNSSPOPILVPELKRSLPALROLPLPDSNAFTQSQOQPPPTTS 653
Qy 284 SDMWIATCTTVTTSPTVTVSSQSEKSIIVSGATAPRTVS 327
Db 654 Q-----ATTALTAIVLVSSVQRTAGKTAATVTSALOPPVLSL 690

RESULT 13
US-09-724-676-63926
; Sequence 63926, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63926
; LENGTH: 807
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-63926

Query Match
Best Local Similarity 38.4%; Score 491; DB 5; Length 807;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

Qy 13 PPKVSSG-----PRLPAPQIVAVKAPNTTTFPANLQLPCTVLIKSNSGPLM 61
Db 414 PTATTSGRATLTPVLAPRLPQ-----PONPTNIO-----NFQLPQGMVLVRSENGQL 464
Qy 62 LVSPQQTIVTR-----AETTSNITSRPVAVPANPQTVKICITVPNSSQILKKVAVTPVKKL 115
Db 465 MI-PQALAQMOAHAPQTTMAPRATPTTSAPPVQISTVQAPCTPIIAR-QVTP----- 518
Qy 116 AOGITVTVTPKPSVSVAVPTSVTVTPGKPLNTVT--TLKPSIGASS-----TPS 168
Db 519 -----TTIIKQV-----SQAQTTVQPSATLQSPGVQPOLVLGGAQATASIGTATAVQGTGPQ 571
Qy 169 NE-PLKAKENSAVOINLSPTMLENVKCKNFMLILKLACSGSGSPMGONVKKLVEQL 227

```

```

; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63926
; LENGTH: 807
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-63926

Query Match
Best Local Similarity 38.4%; Score 491; DB 5; Length 807;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

Qy 13 PPKVSSG-----PRLPAPQIVAVKAPNTTTFPANLQLPCTVLIKSNSGPLM 61
Db 414 PTATTSGRATLTPVLAPRLPQ-----PONPTNIO-----NFQLPQGMVLVRSENGQL 464
Qy 62 LVSPQQTIVTR-----AETTSNITSRPVAVPANPQTVKICITVPNSSQILKKVAVTPVKKL 115
Db 465 MI-PQALAQMOAHAPQTTMAPRATPTTSAPPVQISTVQAPCTPIIAR-QVTP----- 518
Qy 116 AOGITVTVTPKPSVSVAVPTSVTVTPGKPLNTVT--TLKPSIGASS-----TPS 168
Db 519 -----TTIIKQV-----SQAQTTVQPSATLQSPGVQPOLVLGGAQATASIGTATAVQGTGPQ 571
Qy 169 NE-PLKAKENSAVOINLSPTMLENVKCKNFMLILKLACSGSGSPMGONVKKLVEQL 227

```

```

RESULT 9
US-09-724-676-63923
: Sequence 63923; Application US/09724676
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724,676
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ. ID NOS: 97222
: SOFTWARE: PatentIn Version 3.2
: SEQ ID NO 63923
: LENGTH: 775
: TYPE: PR1
: ORGANISM: Homo sapiens
US-09-724-676-63923

```

```

RESULT 11
US-09-724-676-63921
; Sequence 63921, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: CompuGen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 CompuGen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63921
; LENGTH: 805
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-63921

Query Match      27.7%; Score 491; DB 5; Length 805;
Best Local Similarity 38.4%; Pred. No. 5.7e-21;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14,

Oy      13 PKVSSG-----PRLPAPQIVANKANTTTIOEPALQLPQGVLLKSNGLM 61
      |::||| | | | | | | | | | | | | | | | | | | | | | | | |
Db      362 PTAATSGIRATITPPVLAIRLPQ-----PONPTNIQ--NFOLPGKAVLRSENGQL 432

```


XX Dikstein R, Tjian R;
XX WPI; 1998-109818/10.
DR P-PSDB; AAW31494.
XX
XX DNA encoding human tata-binding protein associated factor - for
XX producing recombinant protein
XX
XX Claim 7; Col 11-18; 12pp; English.
XX
CC This cDNA sequence encodes a human tata-binding protein associated
CC factor, hTAFII105, isolated from Daudi cell nuclear extracts. Tightly
CC associated subunits (TAF's) are components of the transcription factor
CC TFIID and are thought to mediate transcriptional activation. This encoded
CC protein may be produced recombinantly from transformed host cells or
CC purified from human cells. hTAFII105 specific binding agents such as
CC specific antibodies could be used for diagnosis (e.g. genetic
CC hybridisation screens for hTAFII105 transcripts), therapy (e.g. gene
CC therapy to modulate hTAFII105 gene expression) and in the
CC biopharmaceutical industry (e.g. as immunogens, reagents for isolating B
CC cell specific activators or other transcriptional regulators).
XX
SQ Sequence 2556 BP; 797 A; 614 C; 514 G; 631 T; 0 other;

Query Match 99.1%; Score 2534; DB 19; Length 2556;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2556; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1 GGGACCTGGTGACCAAAAGTGGCTCGGTCAGCGCCCTCCCTAAAGTCAGCAGCGGCCCT 60
DB 1 GGGACCTGGTGACCAAAAGTGGCTCGGTCAGCGCCCTCCCTAAAGTCAGCAGCGGCCCT 60

QY 61 AGGTCGCTGCTCCCTCAGATAGTGGCCGTGAAGCCCCCAACACACAGCAATFCCAGTTT 120
DB 61 AGGTCGCTGCTCCCTCAGATAGTGGCCGTGAAGCCCCCAACACACAGCAATFCCAGTTT 120

QY 121 CCTGCTAAATTTGAGCTTCTCCAGGAACCGTTTTCATTAAAGTAACAGTGTCCGTTG 180
DB 121 CCTGCTAAATTTGAGCTTCTCCAGGAACCGTTTTCATTAAAGTAACAGTGTCCGTTG 180

QY 181 ATGTGGTATCTCCTCAGCAAACTGTAAAGAGCGGAGACACCAAGTAACATTAACCTCA 240
DB 181 ATGTGGTATCTCCTCAGCAAACTGTAAAGAGCGGAGACACCAAGTAACATTAACCTCA 240

QY 241 AGGCCAGCAGTACCAGCAATCTCTAAACAGTCAAAATCTGTACAGTGGCGAACTTACG 300
DB 241 AGGCCAGCAGTACCAGCAATCTCTAAACAGTCAAAATCTGTACAGTGGCGAACTTACG 300

QY 301 TCACAAATTAATCAAGAAAGTGGCAGTGACACCTGTGTTAAAAAATTTGGCACAATAGGAAT 360
DB 301 TCACAAATTAATCAAGAAAGTGGCAGTGACACCTGTGTTAAAAAATTTGGCACAATAGGAAT 360

QY 361 ACTGGGTAACCACTGTTCCGAAGCCTTCCCTCAGTACAATCTGTGGCTGTGCCAACCACT 420
DB 361 ACTGGGTAACCACTGTTCCGAAGCCTTCCCTCAGTACAATCTGTGGCTGTGCCAACCACT 420

QY 421 RCTGCTCAGCTACTCTCTGGAAGCCATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 480
DB 421 -GTCGTACAGTACTCTCTGGAAGCCATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 479

QY 481 TTTGGGAGCATCATCACCTCTTCAATGAGCCCAATCTTAAAGCAGAGAATCTCAGCAGC 540
DB 480 TTTGGGAGCATCATCACCTCTTCAATGAGCCCAATCTTAAAGCAGAGAATCTCAGCAGC 539

QY 541 TGTTCAGATTAATCTTCTCCGACAAATGCTAGAAAATGTGAAGAAATGCAAGAACTTCTCT 600
DB 540 TGTTCAGATTAATCTTCTCCGACAAATGCTAGAAAATGTGAAGAAATGCAAGAACTTCTCT 599

QY 601 TCCAAATGTAATAAAGTACGATGTAGTGATACAGTCCCTGAAATGGGGGCAAAATGT 660
DB 600 TCCAAATGTAATAAAGTACGATGTAGTGATACAGTCCCTGAAATGGGGGCAAAATGT 659

QY 661 GAAGAAGCTGGTGGAAACAACCTTTTGGATGCAAAAAATCGAAGCAGAGAAATTTACTAGGAA 720
DB 660 GAAGAAGCTGGTGGAAACAACCTTTTGGATGCAAAAAATCGAAGCAGAGAAATTTACTAGGAA 719

QY 721 ACTGTATGTTGAACCTCAAGTCTTACCTCAGCTCAGCTGCTGTTCCCTTTTCTTAAGAAAG 780
DB 720 ACTGTATGTTGAACCTCAAGTCTTACCTCAGCTCAGCTGCTGTTCCCTTTTCTTAAGAAAG 779

QY 781 CGTGGTTGCTTACGACAACTTCTGCTTAACCTCCAGAGCTTTCATCAGCAATGTGTTC 840
DB 780 CGTGGTTGCTTACGACAACTTCTGCTTAACCTCCAGAGCTTTCATCAGCAATGTGTTC 839

QY 841 GCAGACTTCTAGTGACATGGTCAATGCTACTACTACAAAGTAACTTCTCTCTCTCTCTCT 900
DB 840 GCAGACTTCTAGTGACATGGTCAATGCTACTACTACAAAGTAACTTCTCTCTCTCTCTCT 899

QY 901 GGTGACAACTACAGTCTCTCAAGCCAGTCTGAAAGTCAATTTATGTTTCTCGAGCAAC 960
DB 900 GGTGACAACTACAGTCTCTCAAGCCAGTCTGAAAGTCAATTTATGTTTCTCGAGCAAC 959

QY 961 AGCACCAGAACTGTGTGAGTGCAGTGTGAACCCACTTGTGCTGGTCCAGTGGGACAAA 1020
DB 960 AGCACCAGAACTGTGTGAGTGCAGTGTGAACCCACTTGTGCTGGTCCAGTGGGACAAA 1019

QY 1021 AGCTGAGTGTGACACTTCAATCTGTGGCCCAACTGCTGCAACAGGAGGAACACAGC 1080
DB 1020 AGCTGAGTGTGACACTTCAATCTGTGGCCCAACTGCTGCAACAGGAGGAACACAGC 1079

QY 1081 TGAACCTGGTGTGCTTTCAGACTTCAAAACACCTGTGACATCTGTGCAACACAGTGC 1140
DB 1080 TGAACCTGGTGTGCTTTCAGACTTCAAAACACCTGTGACATCTGTGCAACACAGTGC 1139

QY 1141 CAGGCTCTCACTGCAACCTGAAAAGCCAGTGTCTCTGGAACACAGTAACTGCTCCCT 1200
DB 1140 CAGGCTCTCACTGCAACCTGAAAAGCCAGTGTCTCTGGAACACAGTAACTGCTCCCT 1199

QY 1201 TCCAGCAGTAACCTTTTGGAGAACTTTCAGGTGAGCTATTTGCTTCCATCTGTGAAAC 1260
DB 1200 TCCAGCAGTAACCTTTTGGAGAACTTTCAGGTGAGCTATTTGCTTCCATCTGTGAAAC 1259

QY 1261 TGTGTTTCTCTGCTGGGACCACTCTGCAAGCCCTGTTTATTTGGGACTTCAGTTCAAAT 1320
DB 1260 TGTGTTTCTCTGCTGGGACCACTCTGCAAGCCCTGTTTATTTGGGACTTCAGTTCAAAT 1319

QY 1321 CAAACTTGGCCAGCGGGCCCTGCTCTTTCACAAACAGCTGGGATTTCAACAGGCACTTC 1380
DB 1320 CAAACTTGGCCAGCGGGCCCTGCTCTTTCACAAACAGCTGGGATTTCAACAGGCACTTC 1379

QY 1381 AAGCAGCAACTATTTCTCATTTGTTTTCAGCTAGTTCAGGAGCCCTTCAGGAGGCAATGAAA 1440
DB 1380 AAGCAGCAACTATTTCTCATTTGTTTTCAGCTAGTTCAGGAGCCCTTCAGGAGGCAATGAAA 1439

QY 1441 ACAAGTGACCAAAATTTACATTTCTCAACATTTGACCATTTGAGAAATGTGGACAGAGAC 1500
DB 1440 ACAAGTGACCAAAATTTACATTTCTCAACATTTGACCATTTGAGAAATGTGGACAGAGAC 1499

QY 1501 GATGCCAGTGAACACCAATTAATACCTACTAGTCACTTCTTCCAGCTTCCATTTCTAAGCA 1560
DB 1500 GATGCCAGTGAACACCAATTAATACCTACTAGTCACTTCTTCCAGCTTCCATTTCTAAGCA 1559

QY 1561 AATTACTCTGCTGGAAATAAATTTCTCTCACTTCAAGCATCTCTCTACTCAGAAAATA 1620
DB 1560 AATTACTCTGCTGGAAATAAATTTCTCTCACTTCAAGCATCTCTCTACTCAGAAAATA 1618

QY 1621 GAATAAAGAGATGTAAACATCATGCTTCCGAGATGAGATGACATCAATGATGTGACTT 1680
DB 1619 GAATAAAGAGATGTAAACATCATGCTTCCGAGATGAGATGACATCAATGATGTGACTT 1678

QY 1681 CTATGCCAGGGTCAACCTTAATGAAGAAAATGCTGCTATCTTAGCAACAACTCTGAAT 1740
DB 1679 CTATGCCAGGGTCAACCTTAATGAAGAAAATGCTGCTATCTTAGCAACAACTCTGAAT 1738

QY 1741 TGGTTGGCACACTCATTCTAGTCAATGTAAAGATGAACCAATTTCTTTTATTTGGAGCTCTAC 1800

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2003, 22:02:45 ; Search time 513 Seconds
(without alignments)
11229.263 Million cell updates/sec

Title: US-09-763-909-1

Perfect score: 2558

Sequence: 1 gggaccctgtgaccaaagt.....ctttattacttaccctat 2558

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002.*

1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2534	99.1	2556	19	Human hTAFII105 cd
2	2534	99.1	2556	21	Transcription fact
3	2483.6	97.1	3468	17	DNA encoding novel
4	338.2	13.2	3603	13	Human TATA-binding
5	338.2	13.2	3603	18	TATA-binding prote
6	337.2	13.2	2214	15	TATA-binding prote
7	143.8	5.6	4520	23	Drosophila melanog
8	143.8	5.6	4529	23	Drosophila melanog
9	142.2	5.6	4615	15	TATA-binding prote

10	142.2	5.6	4615	17	AA742210	Drosophila TATA-bi
11	142.2	5.6	4615	18	AA779604	TATA-binding prote
12	135.8	5.3	2307	24	ABK32842	DNA encoding human
13	60	2.3	60	24	ABN33770	Human spliced tran
14	58.6	2.3	9525	23	ABL09422	Drosophila melanog
15	58.6	2.3	10083	23	ABL18476	Drosophila melanog
16	55.8	2.2	486	22	AA775507	Polyglutamine trac
17	54	2.1	8201	21	AAA88864	Human dentin sialo
18	54	2.1	8201	24	ABQ73537	Human dentin sialo
19	46.2	1.8	2439	23	ABL18865	Drosophila melanog
20	46.2	1.8	4439	23	ABL18864	Drosophila melanog
21	42.4	1.7	5120	22	AA846677	DNA sequence of hu
22	41.4	1.6	452	22	AAK79268	Human immune/haema
23	41.4	1.6	583	22	AAK79267	Human immune/haema
24	41	1.6	2169	21	AAA30203	Human RING finger
25	40.8	1.6	85680	21	AAF22299	BAC containing rep
26	39.8	1.6	2832	23	ABL10537	Drosophila melanog
27	39.8	1.6	5438	23	ABL10536	Drosophila melanog
28	39.8	1.6	6378	24	ABQ67027	Human angiogenesis
29	39.8	1.6	6378	24	ABL32176	Human immune syste
30	39.8	1.6	136284	24	ABK83575	Human cDNA differe
31	39.6	1.5	2215	24	ABT99688	Mouse ischaemic co
32	39.6	1.5	2375	21	AAA49881	Human pancreatic p
33	39.6	1.5	10266	17	AA733007	Mouse SKY-related
34	39.6	1.5	14704	13	AAQ20685	PKS 741 insert con
35	39.2	1.5	1935	23	ABL27239	Drosophila melanog
36	39.2	1.5	3263	23	ABL10403	Drosophila melanog
37	39.2	1.5	4374	23	ABL27238	Drosophila melanog
38	39.2	1.5	16962	23	ABL10402	Drosophila melanog
39	39	1.5	6354	24	ABL32588	Human immune syste
40	38.8	1.5	4590	22	AAH24065	Yeast AOD9604-asso
41	38.8	1.5	4829	23	ABL13399	Drosophila melanog
42	38.8	1.5	39746	23	ABL13398	Drosophila melanog
43	38.6	1.5	5101	20	AA13453	Drosophila melanog
44	37.8	1.5	1037	21	AAA59242	Enterococcus faeca
45	37.8	1.5	1159	21	AAA59240	Exons E, C and A o
						An EcoRI fragment

ALIGNMENTS

RESULT 1
AAV02872
ID AAV02872 standard; cDNA; 2556 BP.

XX AC AAV02872;

XX DT 28-APR-1998 (first entry)

XX XX Human hTAFII105 cDNA.

XX DE TATA-binding protein associated binding factor 105; human; activator;
KW hTAFII105; transcription factor; TFIIID; transcriptional activation;
KW antibodies; diagnosis; therapy; biopharmaceutical industry; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 1...2406

XX FT /*tag= a

XX FT /product= hTAFII105

XX FT /note= "human tata binding protein associated
factor 105 partial sequence"

XX XX US5710025-A.

XX XX 20-JAN-1998.

XX XX 02-OCT-1996; 96US-0725012.

XX PR 02-OCT-1996; 96US-0725012.

XX XX (REGC) UNIV CALIFORNIA.

Db 1739 TGGTTGGCACACTCATTCAGTCATGTAAGATGAACCATTTCTTTTATTGGAGCTCTAC 1798
 Qy 1801 AAAAGAGAACTTTAGACATTTGGTAAAAAGCATGACATTTACAGAACTTTAACTCTGATGCTG 1860
 Db 1799 AAAAGAGAACTTTAGACATTTGGTAAAAAGCATGACATTTACAGAACTTTAACTCTGATGCTG 1858
 Qy 1861 TGAACCTTGCATCTCCACAGCAACAGAGCACTACGAGGCTTCTAGAAAACTGACTG 1920
 Db 1859 TGAACCTTGCATCTCCACAGCAACAGAGCACTACGAGGCTTCTAGAAAACTGACTG 1918
 Qy 1921 CAATTGCTCAGCATCGAATGACTACTTACAAGCAAGTGAATAATACATCTGCTAGTG 1980
 Db 1919 CAATTGCTCAGCATCGAATGACTACTTACAAGCAAGTGAATAATACATCTGCTAGTG 1978
 Qy 1981 ATACCAAGTTCAGCTCAAAATTTCTTGAAGAGCTGGATCAATTTGGAGAGCAGAGAAAGG 2040
 Db 1979 ATACCAAGTTCAGCTCAAAATTTCTTGAAGAGCTGGATCAATTTGGAGAGCAGAGAAAGG 2038
 Qy 2041 ATTTGGAGAGAGAAATGTTACTTAAGCAGCAGCAAGAGTCGTTCTAAATAAGAGATC 2100
 Db 2039 ATTTGGAGAGAGAAATGTTACTTAAGCAGCAGCAAGAGTCGTTCTAAATAAGAGATC 2098
 Qy 2101 CAGAACAGCTCAGATTTAAAGCAGAGAAAGCAAGAGTTACAGCAATTTGGAATTCGACAGA 2160
 Db 2099 CAGAACAGCTCAGATTTAAAGCAGAGAAAGCAAGAGTTACAGCAATTTGGAATTCGACAGA 2158
 Qy 2161 TACAGCATAGAGAGCTTAATCTCAGAGCTTTGAGAGCTATTGGACCAAGAGAGAGAC 2220
 Db 2159 TACAGCATAGAGAGCTTAATCTCAGAGCTTTGAGAGCTATTGGACCAAGAGAGAGAC 2218
 Qy 2221 CACTAGATCTGGAAATTCAGGGCTTAAAGACACACCTTCTTGGCTTCGGACATCCAGCC 2280
 Db 2219 CACTAGATCTGGAAATTCAGGGCTTAAAGACACACCTTCTTGGCTTCGGACATCCAGCC 2278
 Qy 2281 TGACAGCACCACCAACAGTTGATCGTCCCAAGAAATCACAGAAATCTGCTCAGGACTTGA 2340
 Db 2279 TGACAGCACCACCAACAGTTGATCGTCCCAAGAAATCACAGAAATCTGCTCAGGACTTGA 2338
 Qy 2341 TATTTTGTATGGACAGAGAGAGATGAAGTATTCGAGCTCTATACCTGCGCCCTTC 2400
 Db 2339 TATTTTGTATGGACAGAGAGAGATGAAGTATTCGAGCTCTATACCTGCGCCCTTC 2398
 Qy 2401 TGAAGTGACCACTCCACTCTTCCATCCACATCTGCTATTATTACTGCCAAGAGACACA 2460
 Db 2399 TGAAGTGACCACTCCACTCTTCCATCCACATCTGCTATTATTACTGCCAAGAGACACA 2458
 Qy 2461 AAGCATTTGTCAGCTGCTCCTCAAAATTTCAATTTCTGGAAAAATAACACCAACATGAAAGAG 2520
 Db 2459 AAGCATTTGTCAGCTGCTCCTCAAAATTTCAATTTCTGGAAAAATAACACCAACATGAAAGAG 2518
 Qy 2521 CATTTGTTTACGATTAGAACTTTTATTAACCTTTACCTAT 2558
 Db 2519 CATTTGTTTACGATTAGAACTTTTATTAACCTTTACCTAT 2556

RESULT 2

AAZ90465
 ID AAZ90465 standard; DNA: 2556 BP.

XX
 AC AAZ90465;
 XX

DT 06-JUN-2000 (first entry)

XX Transcription factor subunit TAFII105 polypeptide encoding DNA.

DE TATA box-binding protein associated factor II 105; TAFII105; cancer;
 KW transcription factor; apoptosis; cytostatic; immunosuppressive;
 KW antiinflammatory; virucide; antibacterial; ds.

XX Homo sapiens.

OS
 XX
 FH Key Location/Qualifiers

FT CDS 1..2556
 FT /*tag= a
 FT /transl_except= (pos: 2404..2406, aa: Glx)
 FT /transl_except= (pos: 2499..2501, aa: Glx)
 XX /note= "the stop codon is not indicated"
 PN WO200012699-A1.
 XX
 XX 09-MAR-2000.
 XX
 XX 25-AUG-1999; 99WO-IL00464.
 XX
 XX 27-AUG-1998; 98IL-0125971.
 XX
 XX (YEDA) YEDA RES & DEV CO LTD.
 XX
 XX Dikstein R, Yamit-hezi A;
 XX
 XX WPI: 2000-256640/22.
 XX P-PSDB: AAY57279.
 XX
 XX Polypeptide encoding TATA box binding protein associated factor II 105
 XX useful for treating e.g. cancers and inducing apoptosis has a dominant
 XX negative effect on the normal biological activity of the binding
 XX protein -
 XX
 XX Claim 12; Fig 1; 48pp; English.
 XX
 XX This DNA encodes a polypeptide comprising a (modified) fragment (I) of
 XX a TATA box-binding protein associated factor II 105 (TAFII105). A
 XX pharmaceutical composition comprising (I) or the polynucleotide or an
 XX inhibitor or antagonist of (I) is useful for treating cancers and
 XX inducing apoptosis in pathological cells. The composition is also useful
 XX for treating autoimmune diseases, inflammatory processes and viral or
 XX bacterial infections.
 XX
 XX Sequence 2556 BP; 797 A; 614 C; 514 G; 631 T; 0 other;

Query Match 99.1%; Score 2534; DB 21; Length 2556;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2556; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
 Qy 1 GGGACCCCTGGTGACCAAAAGTGGCTCCGTCAGCGGCCCTCTCTAAAGTCAGCAGCGGCCCT 60
 Db 1 GGGACCCCTGGTGACCAAAAGTGGCTCCGTCAGCGGCCCTCTCTAAAGTCAGCAGCGGCCCT 60
 Qy 61 AGGCTGCCTGCTCCTCAGATAGTCGCGTGAAGGCCCCCAACACCCAGCAATCCAGTTT 120
 Db 61 AGGCTGCCTGCTCCTCAGATAGTCGCGTGAAGGCCCCCAACACCCAGCAATCCAGTTT 120
 Qy 121 CTGCTAAATTTGCAGCTTCTCCAGGAACCGTTTTGATTTAAAGTAAACAGTGGTCCGTTG 180
 Db 121 CTGCTAAATTTGCAGCTTCTCCAGGAACCGTTTTGATTTAAAGTAAACAGTGGTCCGTTG 180
 Qy 181 ATGTTGGTATCTCTCAGCAAAACTGTACAGAGCGCGAGACCAAGTAACATAACCTCA 240
 Db 181 ATGTTGGTATCTCTCAGCAAAACTGTACAGAGCGCGAGACCAAGTAACATAACCTCA 240
 Qy 241 AGGCCAGCAGTACCAGCAATCTCTCAACAGTCAAAATCTGTACAGTCCGCAACTCTAGC 300
 Db 241 AGGCCAGCAGTACCAGCAATCTCTCAACAGTCAAAATCTGTACAGTCCGCAACTCTAGC 300
 Qy 301 TCACAAATTAATCAAGAAAGTGGCAGTGACACCTGTTAAAAAATTTGGCAAAATAGGAAC 360
 Db 301 TCACAAATTAATCAAGAAAGTGGCAGTGACACCTGTTAAAAAATTTGGCAAAATAGGAAC 360
 Qy 361 ACTGTGGTAACCACTGTTCCGAGGCTTCCCTCAGTACAAATCTGTGGCTGTGCCAACCACT 420
 Db 361 ACTGTGGTAACCACTGTTCCGAGGCTTCCCTCAGTACAAATCTGTGGCTGTGCCAACCACT 420
 Qy 421 RGTGCTCAGTACTCTCTGGAAGCCATTTGAATCTGTAACTACCTGTAACCTTCAAG 480
 Db 421 -GTGCTCAGCTTACTCTCTGGAAGCCATTTGAATCTGTAACTACCTTCAAG 479

QY 481 TTTGGAGCATCACTCCCTTCAATGAGCCCAATCTTAAGCAGAGAACTCAGCAGC 540
DB TTTGGAGCATCACTCCCTTCAATGAGCCCAATCTTAAGCAGAGAACTCAGCAGC 539
QY 541 TGTTCAGATTAATCTTTCCGCAATGCTAGAAATGTGAAGAAATGCAAGAACTTCCT 600
DB TGTTCAGATTAATCTTTCCGCAATGCTAGAAATGTGAAGAAATGCAAGAACTTCCT 599
QY 601 TGCATGTTAAATAACTAGCATGTAGTGATCACAGTCCCTGAAATGGGCAAAATGT 660
DB TGCATGTTAAATAACTAGCATGTAGTGATCACAGTCCCTGAAATGGGCAAAATGT 659
QY 661 GAAGAACTGGTGGAAACAATTTTGGATGCAAAATCGAAGCAGAGAAATTTACTAGGAA 720
DB GAAGAACTGGTGGAAACAATTTTGGATGCAAAATCGAAGCAGAGAAATTTACTAGGAA 719
QY 721 ACTGTATGTGAACCTCAAGTCTTCACCTCAGGCTCACTGGTTCCTTTCTTAAGAAAG 780
DB ACTGTATGTGAACCTCAAGTCTTCACCTCAGGCTCACTGGTTCCTTTCTTAAGAAAG 779
QY 781 CGTGGTTCCTTACGACAACTTCTGCCTTAACCTCCAGAGCTTCATCCAGCAATGTGTCA 840
DB CGTGGTTCCTTACGACAACTTCTGCCTTAACCTCCAGAGCTTCATCCAGCAATGTGTCA 839
QY 841 GCAGACTTCTAGTGACATGGTCAATGTGCTACCTGTACTACAACAGTAACAACCTTCCTGT 900
DB GCAGACTTCTAGTGACATGGTCAATGTGCTACCTGTACTACAACAGTAACAACCTTCCTGT 899
QY 901 GGTGACAACTACAGTTCCTCAAGCCAGTCTGAAAGTCAATATTTGTTCTTGGAGCAAC 960
DB GGTGACAACTACAGTTCCTCAAGCCAGTCTGAAAGTCAATATTTGTTCTTGGAGCAAC 959
QY 961 AGCACCAGAACTGTGTCACTGCAAACTTTGAACCCACTTGTGTCCAGTGGGACAAA 1020
DB AGCACCAGAACTGTGTCACTGCAAACTTTGAACCCACTTGTGTCCAGTGGGACAAA 1019
QY 1021 AGTGGAGTTGTGACACTTTCATCTGTGGGCCAACTGTGTGCAACAGGAGGAAACACAGC 1080
DB AGTGGAGTTGTGACACTTTCATCTGTGGGCCAACTGTGTGCAACAGGAGGAAACACAGC 1079
QY 1081 TGGAACTGTTGCTTTCACACTTCAAAACCACTGTGTGACATCTGTGGCAACACAGTGAC 1140
DB TGGAACTGTTGCTTTCACACTTCAAAACCACTGTGTGACATCTGTGGCAACACAGTGAC 1139
QY 1141 CAGGCTTCACCTGCAACCTGAAAGCCAGTGTCTCTGGAACAGCAGTAACTGTCCCT 1200
DB CAGGCTTCACCTGCAACCTGAAAGCCAGTGTCTCTGGAACAGCAGTAACTGTCCCT 1199
QY 1201 TCCAGCAGTAACCTTTGGAGAACTTCAGGTGACGTAATTTGTCTTCCATCTGTGAAACC 1260
DB TCCAGCAGTAACCTTTGGAGAACTTCAGGTGACGTAATTTGTCTTCCATCTGTGAAACC 1259
QY 1261 TGTGTTCTTCTCTGCTGGCAACACATCTGCAAGCTGTATTGGGACTCCAGTCCAAAT 1320
DB TGTGTTCTTCTCTGCTGGCAACACATCTGCAAGCTGTATTGGGACTCCAGTCCAAAT 1319
QY 1321 CAAACTTGGCCAGCGGGCCCTGTCTTTTCAACACAGCTGGGATTCGAAGCAGGTTC 1380
DB CAAACTTGGCCAGCGGGCCCTGTCTTTTCAACACAGCTGGGATTCGAAGCAGGTTC 1379
QY 1381 AAGCAAGCAACTATTCCTATGTTTCACTAGTTTCAAGAGCTTCAGAGGCAATGAAAA 1440
DB AAGCAAGCAACTATTCCTATGTTTCACTAGTTTCAAGAGCTTCAGAGGCAATGAAAA 1439
QY 1441 ACAAGTGACCAAAATTCACATTCCTCAACATTTGACATTCAGAAATGTGGCAGAGAC 1500
DB ACAAGTGACCAAAATTCACATTCCTCAACATTTGACATTCAGAAATGTGGCAGAGAC 1499
QY 1501 GATGCCAGTGAACACATAATACCTACTAGTCAGTTTCTCCAGCTTCCATTTCTAAAGCA 1560
DB GATGCCAGTGAACACATAATACCTACTAGTCAGTTTCTCCAGCTTCCATTTCTAAAGCA 1559

QY 1561 AATTACCTCTGCTGGAAATAAAATTTCTGTCACTTCAAGCATCTCTTACTCAGAAAAATA 1620
DB AATTA-CTCTGCTGGAAATAAAATTTCTGTCACTTCAAGCATCTCTTACTCAGAAAAATA 1618
QY 1621 GAATAAAGAGAATGTAAACATCATGCTTCCGAGATGAGGATGACATCAATGTGTGACTT 1680
DB GAATAAAGAGAATGTAAACATCATGCTTCCGAGATGAGGATGACATCAATGTGTGACTT 1678
QY 1681 CTATGTCAGGGTCAACCTTAATGAAGAAATGCCTGCATCTTAGCAACAACTCTGAAT 1740
DB CTATGTCAGGGTCAACCTTAATGAAGAAATGCCTGCATCTTAGCAACAACTCTGAAT 1738
QY 1741 TGGTTGGCACACTCATTCAGTCAATGAAGATGAACCAATTTCTTTTATTTGGAGCTCTAC 1800
DB TGGTTGGCACACTCATTCAGTCAATGAAGATGAACCAATTTCTTTTATTTGGAGCTCTAC 1798
QY 1801 AAAGAGAATCTTAGACATTTGGTAAAGATGACATTTACAGAACTTAACCTGTGAGCTG 1860
DB AAAGAGAATCTTAGACATTTGGTAAAGATGACATTTACAGAACTTAACCTGTGAGCTG 1858
QY 1861 TGAACCTTGATCTCCCAAGCAACACAGAGACTACGAGGCTTCTAGAAAACTGACTG 1920
DB TGAACCTTGATCTCCCAAGCAACACAGAGACTACGAGGCTTCTAGAAAACTGACTG 1918
QY 1921 CAATTGCTCAGCATCGAATGACTTACTTACAAGCAAGTGAATAATTTACATCTGTGAGTG 1980
DB CAATTGCTCAGCATCGAATGACTTACTTACAAGCAAGTGAATAATTTACATCTGTGAGTG 1978
QY 1981 ATACAGGTCACAGCTCAAAATTTCTTGAAGCTGGATCAATTTGGAGAGCAGAGAAAG 2040
DB ATACAGGTCACAGCTCAAAATTTCTTGAAGCTGGATCAATTTGGAGAGCAGAGAAAG 2038
QY 2041 ATTTGAAGAAAGAGAAATGTTTAAAGCAGCCAGAGTCTGTCTTAATAAAGAGATC 2100
DB ATTTGAAGAAAGAGAAATGTTTAAAGCAGCCAGAGTCTGTCTTAATAAAGAGATC 2098
QY 2101 CAGAACGCTGAGATTTAAAGCAGAAAGCCAAAGAGTTACAGCAATTTGGAATTTGCACAGA 2160
DB CAGAACGCTGAGATTTAAAGCAGAAAGCCAAAGAGTTACAGCAATTTGGAATTTGCACAGA 2158
QY 2161 TACAGCATAGAGACGCTAATCTCACAGCTTTCAGACTATTGGACCAAGGAAGAGAC 2220
DB TACAGCATAGAGACGCTAATCTCACAGCTTTCAGACTATTGGACCAAGGAAGAGAC 2218
QY 2221 CACTAGAATCTGGAATTTAGGGCTTAAAGACAAACCTTCTTGTCTTGGGACATCCAGCC 2280
DB CACTAGAATCTGGAATTTAGGGCTTAAAGACAAACCTTCTTGTCTTGGGACATCCAGCC 2278
QY 2281 TGACAGCCACCAACAGTTGCAATCGTCCAAAGATCAGAGAACTCGCTCAGGACTTGA 2340
DB TGACAGCCACCAACAGTTGCAATCGTCCAAAGATCAGAGAACTCGCTCAGGACTTGA 2338
QY 2341 TATTTTGTATGGAACAGGAGGAGATGAAGTATTCTCGAGCTTATACCTGGCCCTTC 2400
DB TATTTTGTATGGAACAGGAGGAGATGAAGTATTCTCGAGCTTATACCTGGCCCTTC 2398
QY 2401 TGAAGTGACACTCCACTCTTCCATCCACATCTTGTCTATTACTGCCAAGAGACACA 2460
DB TGAAGTGACACTCCACTCTTCCATCCACATCTTGTCTATTACTGCCAAGAGACACA 2458
QY 2461 AAGCATTTGTGACTGTCTGAAATTTCAATTTCTGGAAAAATTAACCAACATGAAGAG 2520
DB AAGCATTTGTGACTGTCTGAAATTTCAATTTCTGGAAAAATTAACCAACATGAAGAG 2518
QY 2521 CATTTGTTACGATTAGAACTTTTAACTTTTACCTAT 2558
DB CATTTGTTACGATTAGAACTTTTAACTTTTACCTAT 2556

RESULT 3
AAS73655
ID AAS73655 standard; cDNA; 3468 BP.
XX

QY 1261 TGTGTTTCTCTGCTGGGACCACATCTG-CAAGCCTGTTATTTGGGACTCCAGTTCAAA 1319
 Db 1260 TTTGTTTCTCTGCTGGGACCACATCTGACAAGCCTGTTATTTGGGACTCCAGTTCAAA 1319
 QY 1320 TCAAACTTCCCGACGGGCGCTGCTTTTACACACGAGCTGGGATTTCCAAACAGGCGATT 1379
 Db 1320 TCAAACTTCCCGACGGGCGCTGCTTTTACACACGAGCTGGGATTTCCAAACAGGCGATT 1379
 QY 1380 CAAG-CAAGCAACTATTTCTCATTTGTTTTCACGTAGTTTCAGCAGCCTTCAGGAGGCAATGAA 1438
 Db 1380 CAAGTCAAGCAACTATTTCTCATTTGTTTTCACGTAGTTTCAGCAGCCTTCAGGAGGCAATGAA 1439
 QY 1439 AAACAAGTACACCAATTTTCACATTCCTCAACATTTGACCATTTCAAGAAATGTGGACAGAG 1498
 Db 1440 AAACAAGTACACCAATTTTCACATTCCTCAACATTTGACCATTTCAAGAAATGTGGACAGAG 1499
 QY 1499 ACGATGCCAGTGAACACCAATAACTACTAGTCAAGTTTCCCTCCAGCTTCCATTTCTAAAG 1558
 Db 1500 ACGATGCCAGTGAACACCAATAACTACTAGTCAAGTTTCCCTCCAGCTTCCATTTCTAAAG 1559
 QY 1559 CAATTTACCTCTGCTGGGAAATAAATTTCTGTCACCTTCAAGCATCTCTCTACTCAGAAAAA 1618
 Db 1560 CAATTTA-CTCTGCTGGGAAATAAATTTCTGTCACCTTCAAGCATCTCTCTACTCAGAAAAA 1618
 QY 1619 TAGAATAAAGAGATGTAAACATCATGCTTCCGAGATGAGGATGACATCAATGATGTGAC 1678
 Db 1619 TAGAATAAAGAGATGTAAACATCATGCTTCCGAGATGAGGATGACATCAATGATGTGAC 1678
 QY 1679 TTCTATGGCAGGGTCAACCTTAATGAGAAAAATGCTGCATCTTAGCAACAAACTCTGA 1738
 Db 1679 TTCTATGGCAGGGTCAACCTTAATGAGAAAAATGCTGCATCTTAGCAACAAACTCTGA 1738
 QY 1739 ATTGGTTGGCACACATTCAGTCATGTAAAGATGAACCATTTCTTTTATTGGAGCTCT 1798
 Db 1739 ATTGGTTGGCACACATTCAGTCATGTAAAGATGAACCATTTCTTTTATTGGAGCTCT 1798
 QY 1799 ACAAAAGAGAACTTTAGACATTTGGTAAAGATGACATTAACAGAACTTAACCTCTGATGC 1858
 Db 1799 ACAAAAGAGAACTTTAGACATTTGGTAAAGATGACATTAACAGAACTTAACCTCTGATGC 1858
 QY 1859 TGTGAACCTTGATCTCCCAAGCAACAGGAGGAGCTACGAGGCTTCTAGAAAAAAGTAC 1918
 Db 1859 TGTGAACCTTGATCTCCCAAGCAACAGGAGGAGCTACGAGGCTTCTAGAAAAAAGTAC 1918
 QY 1919 TGCAATTTGCTCAGCATGCAATGACTTACAGGCAAGTGAAGAAATTTACATCTGTGTAG 1978
 Db 1919 TGCAATTTGCTCAGCATGCAATGACTTACAGGCAAGTGAAGAAATTTACATCTGTGTAG 1978
 QY 1979 TGATACCAAGTCAACAGCTCAAAATTTCTTGAAGAGCTGGATCAATTTGGAGAACAGAAA 2038
 Db 1979 TGATACCAAGTCAACAGCTCAAAATTTCTTGAAGAGCTGGATCAATTTGGAGAACAGAAA 2038
 QY 2039 GGATTTGGAAGAAAGAGAAATGTTACTTAAAGCAGCAAGAGTCTGCTTAATAAAGAGA 2098
 Db 2039 GGATTTGGAAGAAAGAGAAATGTTACTTAAAGCAGCAAGAGTCTGCTTAATAAAGAGA 2098
 QY 2099 TCCAGAACAGCTGAGATTAAGCAGAAAGCCAAAGAGTTTACAGCAATTTGGAATTTGCACA 2158
 Db 2099 TCCAGAACAGCTGAGATTAAGCAGAAAGCCAAAGAGTTTACAGCAATTTGGAATTTGCACA 2158
 QY 2159 GATACAGCATAGACAGCTTAATCTCACAGCTCTTCCAGCTATTGGACCAAGGAAGAAG 2218
 Db 2159 GATACAGCATAGACAGCTTAATCTCACAGCTCTTCCAGCTATTGGACCAAGGAAGAAG 2218
 QY 2219 ACCACTAGAAATCTGGAATTTGAGGCTTAAAGACAACTTCTTGTCTTCTGGGACATCCAG 2278
 Db 2219 ACCACTAGAAATCTGGAATTTGAGGCTTAAAGACAACTTCTTGTCTTCTGGGACATCCAG 2278
 QY 2279 CCTGACAGCCCAACAGTTGTCATGCTGCTCAAGAAATTCAGGAAATCTGCCTCAGGACTT 2338
 Db 2279 CCTGACAGCCCAACAGTTGTCATGCTGCTCAAGAAATTCAGGAAATCTGCCTCAGGACTT 2338
 QY 2339 GATATTTTGTATGGAACAGGAGGAGATGAAGTATTTCTCCAGCTCTATACCTGGCCCT 2398

Db 2339 GATATTTTGTATGGAACAGGAGGAGATGAAGTATTTCTCCAGCTCTATACCTGGCCCT 2398
 QY 2399 TCTGAAGTGCACCACTCCACTTCTCCATCCACATCCTTGTCTATTACTGCGCAAGAGACA 2458
 Db 2399 TCTGAAGTGCACCACTCCACTTCTCCATCCACATCCTTGTCTATTACTGCGCAAGAGACA 2458
 QY 2459 CAAAGCATTTGTCACCTGCTGAAATTTCAATTTCTGGAATAA-CACCAACATGAA 2517
 Db 2459 CAAAGCATTTGTCACCTGCTGAAATTTCAATTTCTGGAATAAATATCACCACATGAAT 2518
 QY 2518 GAGCATTTGTCAGATAG-AACTTTTATTAACTCTTTACCTAT 2558
 Db 2519 GAGCATTTGTCAGATAGAAACTTTTAACTCTTTACCTAT 2560

RESULT 4
 AAT42217
 ID AAT42217 standard; cDNA; 3603 BP.
 XX
 AC AAT42217;
 XX
 DT 27-JAN-1997 (first entry)
 XX
 DE Human TATA-binding protein associated factor hTAFII130 gene.
 KW Human; TATA-binding protein; TBP associated factor; TFIID; holoenzyme;
 KW RNA polymerase II; transcription; messenger RNA; nuclear fraction;
 KW lambda-gt11; expression library; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 CDS 1..2214
 FT /*tag= a
 FT /product= human TAFII130
 XX
 PN US5534410-A.
 XX
 PD 09-JUL-1996.
 XX
 PF 28-JAN-1993; 93US-0013412.
 XX
 PR 28-JAN-1994; 94US-0188582.
 PR 28-JAN-1993; 93US-0013412.
 PR 30-JUN-1993; 93US-0087119.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
 PI Tjian R, Wang E, Weinzierl ROJ;
 XX
 DR WPI; 1996-333245/33.
 DR P-PSDB; AAW06084.
 XX
 PT Screen for cpds. that bind human TATA-binding protein associated
 PT factor - by testing ability to bind to polypeptide fragments of the
 PT factor, useful as (ant)agonists of transcription factors involved in
 PT disease.
 XX
 PS Examples; Column 105-112; 86pp; English.
 CC
 CC This is the nucleotide sequence encoding the human TATA-binding protein
 CC (TBP) associated factor (TAF) designated TAFII130. The protein is a
 CC component of the TFIID fraction required for reconstituting RNA
 CC polymerase II in vitro transcription activity. The encoded protein has
 CC an estimated mol. wt. of 130 kD by SDS-PAGE.
 CC The invention relates to purified proteins involved in transcription
 CC by RNA polymerase II, the RNA polymerase which transcribes messenger
 CC RNA. RNA polymerase II transcription proceeds in vitro upon addition
 CC of several nuclear fractions designated TFIIA, B, D, E, F, H, I and J
 CC to RNA polymerase II holoenzyme. Fraction TFIID has been shown to
 CC contain a TBP and other TAFs. Purification of TFIID and separation of

CC its components reveals 7 proteins ranging in size from 30-250 kD.
 CC Serum raised against the TFIID fraction allowed cloning of the corresp.
 CC genes from lambda-gt11 expression libraries.

XX SQ Sequence 3603 BP; 969 A; 961 C; 830 G; 843 T; 0 other;

Query Match 13.2%; Score 338.2; DB 17; Length 3603;
 Best Local Similarity 63.9%; Pred. No. 1.6e-84;
 Matches 536; Conservative 0; Mismatches 288; Indels 15; Gaps 1;

QY 1586 TCTGTCATCTCAAGCATCTCTACTCAGAAAAATAGATAAAGAGAAATGTAACATCATG 1645
 DB 1377 TGTCTCGCACAGCAGCTGCTGCACAGAAAAATAAACTCAAGAGCGCTGGGGAGGTT 1436
 QY 1646 CTTCCGAGATGAGGATGACATCAATGATGCTCTTATGCGAGGGGTCAACCTTAATGA 1705
 DB 1437 GTTTCCGGGAGCATGATGACATTAATGATGTTGTCATGCTGCAGTAAACTTGTGAGA 1496
 QY 1706 AGAAATGCGCTGTCATCTAGCAACAACTCTGAATGTTGGTGGCACATCTTCAGTCTG 1765
 DB 1497 AGAAGTGCAGAAATATTAGCCAGCAACTCTGAATGTTGGTGGCACGCTTAACGGGCT 1556
 QY 1766 TAAAGATGAACCACTTCTTTTATTGGAGCTCTACAAAAGAGAAATCTTAGACATTTGTA 1825
 DB 1557 TAAAGATGAACCACTTCTCTCCCAAGCGCTTTCAGAGAGAAATATTAGAAATAGTAA 1616
 QY 1826 AAGCATGACATTAACAGAACTTAACCTGATGCTGTGAATGTTGATCCCAAGCAACACA 1885
 DB 1617 AAAACATGGTATAACGGAATTTACATCCAGATGTAAGTATGATATCATGTCACAGCA 1676
 QY 1886 GGAACGACTAGGAGCGCTCTAGAAAACCTGACTGCAATGCTCAGCATCGAATGACTAC 1945
 DB 1677 ACAAGCTACAGAACTCTGTAGAGAAATATTCAGAAACAGCTCAGCAGAGAACTTTTC 1736
 QY 1946 TTACAAGCAAGTGAATTAACATCTCTGTGTGATACACAGGTCAAGCTCAAAATTTCT 2005
 DB 1737 TTACAAGTATGACGACAGATGATGAGCGGAGTGACGTGCGGSCACAGCTCAAGTTT 1796
 QY 2006 TGAAGACTGATCAATTTGGAGAGCAGAGAAAGGATTTGGAAGAAAGAGAAATGTTACT 2065
 DB 1797 TGAACAGCTTGTATCAAAATCGAAAAGCAGAGAGAGATGAGCAGGCGGAGATCTTAT 1856
 QY 2066 TAAGGAGCCAGAGTGTCTTAATAAAGAGATCCAGAACAGCTGAGATTAAAGCAGAA 2125
 DB 1857 GAGGCGAGCAAGTCTCGGTCAAGACAGAAAGATCCAGAACAGTAAAGGCTGAACAGAA 1916
 QY 2126 AGCCAAAGAGTTACAGCAATTTGGAACCTTGCCAGATACAGCATAGACAGCTTAATCTCAC 2185
 DB 1917 GGCAAGAGATGACGACCAACAGGAATGCGCAATGAGACAGCGGCGGACCCACCTCAC 1976
 QY 2186 AGCTTTGACGCTATTGACCAAGAGAAAGAGACACCACTAGA-----ATC 2230
 DB 1977 AGCACTAGCAGCATCGGGCCCGCAGCAAAAAGAGAGAAAGTGGAGTGTCCGGGCGGCGTC 2036
 QY 2231 TGGAAATGAGGCTTAAAGACACCTTCTGTCTTCTGGGACATCCAGCTGACAGCCAC 2290
 DB 2037 AGGAGCAGAGGTTGGGCGCCGCTCAGTGTGTCAGGAGGAGTGTGCGGAGTGTGCGGA 2096
 QY 2291 CAACAGTTGCTGCTGCAAGAAATCAGAGAAATCTGCTCAGGAGCTTGATATTTGTTAT 2350
 DB 2097 CAGACAGTTTCCGCGACAAAGATCAGCGGGTCAACCTCAGGAGCTCATATTTTGT 2156
 QY 2351 GGAACAGAGAGGAGATGAGTATTTCTGAGCTCTATACCTGCGCCCTCTCTGAAGTAC 2409
 DB 2157 AGAAATGAACGTGAGACAGCCATTTCACTGCTCTCAAAAAGCATTTCTTAAAGTAC 2215

RESULT 5
 AAT79595
 ID AAT79595 standard; cdna; 3603 BP.
 XX
 AC AAT79595;
 XX

DT 08-OCT-1997 (first entry)
 DE TATA-binding protein associated factor, hTAFII30 cDNA.
 XX
 KW TATA-binding protein associated factor; TAF; nuclear protein;
 KW RNA polymerase transcription; TATA-binding protein; TBP;
 KW initiation; ds.
 XX Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..2214
 FT /*tag= a
 FT /note= "no start codon"

PN US5637686-A.
 XX
 PD 10-JUN-1997.
 XX
 PF 28-JAN-1993; 93US-0013412.
 XX
 PR 28-JAN-1994; 94US-0188582.
 PR 28-JAN-1993; 93US-0013412.
 PR 30-JUN-1993; 93US-0087119.
 PR 09-MAY-1996; 96US-0846715.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Comai L, Dynlact BD, Hoey T, Ruppert S, Tanese N;
 PI Tjian R, Wang E, Weinzierl ROJ;
 XX
 DR WPI; 1997-319113/29.
 DR P-PSDB; AAW25019.

XX Nucleic acids encoding human TATA-binding protein associated factor
 PT (TAF) peptide(s) - for production of recombinant peptide(s), used
 PT for modulating transcription of TAFs

PS Claim 11; Column 105-112; 86pp; English.

XX AAT79595 encodes human TATA-binding protein associated factor (TAF)
 CC polypeptide, hTAFII130 (mol. weight 130kD). TAF peptides derived
 CC from hTAFII30 alpha, hTAFII30 beta, hTAFII40, hTAFII70, hTAFII100,
 CC hTAFII130, hTAFII250, hTAFII48 and hTAFII10 and nucleic acids encoding
 CC them, are used to modulate transcription, including transcription
 CC initiation. TAFs are nuclear proteins involved in RNA polymerase I,
 CC II and III transcription. The peptides act by binding to a different
 CC TAF, an activator, or TBP (TATA-binding protein) or competitively
 CC inhibiting association of a TAF domain with another compound, typically
 CC a protein like TBP or another TAF, an activator, or DNA.

XX SQ Sequence 3603 BP; 969 A; 961 C; 830 G; 843 T; 0 other;

Query Match 13.2%; Score 338.2; DB 18; Length 3603;
 Best Local Similarity 63.9%; Pred. No. 1.6e-84;
 Matches 536; Conservative 0; Mismatches 288; Indels 15; Gaps 1;

QY 1586 TCTGTCATCTCAAGCATCTCTACTCAGAAAAATAGATAAAGAGAAATGTAACATCATG 1645
 DB 1377 TGTCTCGCACAGCAGCTGCTGCACAGAAAAATAAACTCAAGAGCGCTGGGGAGGTT 1436
 QY 1646 CTTCCGAGATGAGGATGACATCAATGATGCTCTTATGCGAGGGGTCAACCTTAATGA 1705
 DB 1437 GTTTCCGGGAGCATGATGACATTAATGATGTTGTCATGCTGCAGTAAACTTGTGAGA 1496
 QY 1706 AGAAATGCGCTGTCATCTAGCAACAACTCTGAATGTTGGTGGCACATCTTCAGTCTG 1765
 DB 1497 AGAAGTGCAGAAATATTAGCCAGCAACTCTGAATGTTGGTGGCACGCTTAACGGGCT 1556
 QY 1766 TAAAGATGAACCACTTCTTTTATTGGAGCTCTACAAAAGAGAAATCTTAGACATTTGTA 1825
 DB 1557 TAAAGATGAACCACTTCTCTCCCAAGCGCTTTCAGAGAGAAATATTAGAAATAGTAA 1616

QY 1826 AAAGCATGACATTACAGAACTTAACCTCTGATGCTGTGAACCTTGATCTCCCAAGCAACACA 1885
 DB 1617 AAAACATGATAACCGAATATACATCAGATGTAGTAACTTATGATACATCCACGCA 1676
 QY 1886 GGAACCACTACGAGGCTTCTAGAAAACCTGACTGCAATTTGCTCAGCATCGAATGACTAC 1945
 DB 1677 ACNAGGCTACAGATCTTGTAGANNAATATCAGANACAGCTCAGCAGAAAGCACTTTTC 1736
 QY 1946 TTACAAGCAAGTGAATATACATCTGTGTATGATACCAAGTTCACAGCTCAAAATTTCT 2005
 DB 1737 TTACAAGGATGACGACAGATATGAGCAGCGAGTGTGCTCCGGGCACACCTCAAGTTTTC 1796
 QY 2006 TGAAGAGCTGGATCAATTCGAGACAGAGAAAGATTTGGAGAAAGAAATGTTACT 2065
 DB 1797 TGAACAGCTTGATCAATCGAAAGCAGAGAGGATGACGAGGAGCGGAGATCTGAT 1856
 QY 2066 TAAGGAGCAGCAAGAGTCTGTCTTAATAAGCAAGATCCAGAACAGCTCAGATTAAAGCAGAA 2125
 DB 1857 GAGGCGAGCAAGTCTCGGTCAAGACAGAGATCCAGAACAGTTAAGGCTGAACACAGAA 1916
 QY 2126 AGCCAAAGAGTTACAGAACTTGAACATTCGACAGATACAGCATAGAGACGCTAATCTCAC 2185
 DB 1917 GCAAAAGGAGATGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1976
 QY 2186 AGCTCTTCAGCTATTTGGACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2230
 DB 1977 AGCACTAGCAGGATCGGGCCCGAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2036
 QY 2231 TGAAGATTGAGGCTTAAAGACAACTTCTTCTGTCGAGACATCCAGCTCAGACGCCAC 2290
 DB 2037 AGGAGCAGAGGGTGGGCGCGGCTCAGTGTCTCCAGGACCTCGGGTGGCGGACCC 2096
 QY 2291 CAACAGATTGCTGCTCCAGAAATCACGAGAACTCTGCTCAGGAGCTTGTATTTTGTAT 2350
 DB 2097 CAGACAGTTCCGCGACAAAGAAATCACGGGGTCAACCTCAGGACCTCATATTTTGTAT 2156
 QY 2351 GGAACAGGAGGAGAGATGAGTATTCGAGCTCTATACCTGGGCTTCTGAGGTGAC 2409
 DB 2157 AGAAAATGAACGTGAGACAAGCCATTCAGTGTCTACAAAGCAATTCCTTAAAGTGAC 2215

RESULT 6

NAQ070731

ID NAQ070731 standard; cDNA: 2214 BP.

AC AAQ070731;

XX AAQ070731;

XX 23-MAR-1995 (first entry)

DE TATA-binding protein-associated factor hTAFII130 cDNA.

XX TATA-binding protein associated factor; hTAFII130; ss; screening;

KW diagnostic; therapeutic; gene transcription regulation.

XX Homo sapiens.

XX Key Location/Qualifiers

FT 1..2214

FT /*tag= a

XX W09417087-A.

XX 04-AUG-1994.

XX 28-JAN-1994; 94WO-US01114.

XX 28-JAN-1993; 93US-0013412.

XX 30-JUN-1993; 93US-0087119.

XX (REGC) UNIV CALIFORNIA.

XX Comai L, Dynlact BD, Hoey T, Ruppert S, Tanese N;

PI Tjlan R, Wang E, Weinzierl ROJ;

XX

DR WPI; 1994-264019/32.

DR P-PSDB; AAR56494.

XX

PT TATA-binding protein associated protein factors - and

PT corresponding nucleotide sequence and deriv. antibodies, useful

PT in screening, diagnostics and therapeutics

XX

PS Disclosure; Page 137-141; 180pp; English.

XX

CC The TATA-binding protein associated factor hTAFII130 (including
 CC specific antibodies and fusion products) are used in drug screening,
 CC diagnostics and therapeutics. They are used in the development of
 CC specific biochemical assays for screening compounds that agonise or
 CC antagonise selected transcription factors involved in regulating
 CC gene expression associated with human pathology.

XX

SQ Sequence 2214 BP; 563 A; 691 C; 600 G; 360 T; 0 other;

Query Match

Best Local Similarity 13.2%; Score 337.2; DB 15; Length 2214;

Matches 535; Conservative 0; Mismatches 288; Indels 15; Gaps 1;

QY

1586 TCTGTCACCTCAAGCATCTCTACTCAGAAAATAGATAAAGAGAAATGTAAACATCATG 1645

DB 1377 TGTCTCGGCACAAAGCAGCTGCTGCACAGAAAATAAACAACAAGGAGCTGGGGAGGTTT 1436

QY

1646 CTTCCGAGATGAGGATGACATCAATCATCTGACTTCTATGGCAGGGGTCAACCTTAATGA 1705

DB 1437 GTTTCGGGACGATGATGATTAATGATGTTGCATCGATGGCTGGAGTAAATTTGTCAGA 1496

QY

1706 AGAAATGCTGCATCTTAGCAACAAACTCTGAATTTGGTGGCAGCTCATCTCAGTCATG 1765

DB 1497 AGAAGTGCAGAAATATAGCCAGCACTCTGAATTTGGTGGCAGCTCAACGGGTCCTG 1556

QY

1766 TAAAGATGAACCATTTCTTTTATTTGGAGCTCTACAAAAGAGAAATTTAGACATTTGTTAA 1825

DB 1557 TAAAGATGAACCTTCTCTCCAAAGCGCTTTGCAGAGAAAGATATTAGAAATAGGTAA 1616

QY

1826 AAAGATGACATTACAGAACTTAACCTCTGATGCTGTGAACCTTGAATCTCCCAAGCAACACA 1885

DB 1617 AAACATGTTATACGGAATTTACATCCAGATGTAGTAACTTATGTATCATCTGCCACGA 1676

QY

1886 GGAAGCTACGAGGCTTCTAGAAAACCTGACTGCAATTTGCTCAGCATCGAATGACTAC 1945

DB 1677 ACAAGGCTACAGAACTTTGTAGAGAAATATCAGAAACAGCTCAGCAAGAACTTTTC 1736

QY

1946 TTACAAGGCAAGTGAATATACATCTGTGTATGATACAGCTCAGCTCAAAATTTCT 2005

DB 1737 TTACAAGGATGACGACAGATATGAGCAGCGAGTGTGACGCTCCGGGCACAGCTCAAGTTTTC 1796

QY

2006 TCAAAAGCTTGGATCAATTTGGAGAACAGAGAAAGGATTTGGAGAAAGAGAAATGTTACT 2065

DB 1797 TGAACAGCTTGATCAATTCGAAAGCAGAGGAGGATGAGCAGGAGGAGGAGATCTGAT 1856

QY

2066 TAAGCAGCAGAGTCTGTTCTTAATAAGAAAGATCCAGAACAGCTGAGATTTAAAGCAGAA 2125

DB 1857 GAGGCGAGCAAGTCTCGGTCAAGACAAAGATCCAGAAACAGTAAAGCTGNAACAGAA 1916

QY

2126 AGCCAAAGAGTTACAGAACTTGAACATTCGACAGATACAGCATAGAGACGCTAATCTCAC 2185

DB 1917 GCAAAAGGAGATGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1976

QY

2186 AGCTCTTCAGCTATTTGGACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2230

DB 1977 AGCACTAGCAGGATCGGGCCCGAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2036

QY

2231 TGAAGATTGAGGCTTAAAGACAACTTCTTCTTCTGGGACATCCAGCTCAGACGCCAC 2290

DB 2037 AGGAGCAGAGGGTTCGGGCGCGGCTCAGTGTGTCCAGGACGCTCGGGTGTTCGGAACCCC 2096

QY

2291 CAACAGTTGCTATCGTCCAGAAATCAGAGAAATCTGCCTCAGGAGCTTGTATTTTGTAT 2350

Db 2097 CAGACAGTTCACGGCACAAGAAATACGCGGTCAACCTCAGGGAACCTCATATTTCTTT 2156
 QY 2351 GGAACAGGACGGGAGATGAAGTATTCAGAGCTCTATACCTGCGCCCTCTGAAGTGA 2408
 Db 2157 AGAAATGAACGTGAGAGCAAGCCATTCAGTCTGCTCTACAAAGCATTCCTTAAAGTGA 2214

RESULT 7
 ABL09423
 ID ABL09423 standard; cDNA; 4520 BP.
 AC ABL09423;
 XX 26-MAR-2002 (first entry)
 DT Drosophila melanogaster expressed polynucleotide SEQ ID NO 22751.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX Drosophila melanogaster.
 OS Drosophila melanogaster.
 XX WO200171042-A2.
 XX 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR P-PSDB; ABB65320.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX Claim 1; SEQ ID NO 22751; 2lpp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 4520 BP; 1350 A; 1102 C; 1071 G; 917 T; 0 other;
 SQ

Query Match 5.6%; Score 143.8; DB 23; Length 4520;
 Best Local Similarity 50.8%; Pred. No. 1.4e-29;
 Matches 396; Conservative 0; Mismatches 377; Indels 6; Gaps 2;

QY 1653 GATGAGGATGACATCATGATGTGACATTCATGCGAGGGTCAACCTTAATGAAGAAAT 1712
 Db 2491 GGTGACGATGATATCAACGATGTGCGCCATGCGAGGTGTTAACTTGGCGGAGGTCG 2550
 QY 1713 GCCTGCACTTAGCAACAACTCTGAATGTTGGCACATCATCTAGTCAATGTAAGAT 1772
 Db 2551 CAGGCAATCTCGGCTGTA---CCGAACACATCGCGACGAGATTCGATCTGCAAGAT 2607
 QY 1773 GAACCAATTTCTTTTATTTGGAGCTCTACAAAAGAGATCTTAGACATTTGGTAAAAGAT 1832
 Db 2608 GAGTATTTCTCAATCTCCCTCGCTGCAAGCTAGATACGGGCAATTAATCTCGAGCG 2667

RESULT 8
 ABL18477
 ID ABL18477 standard; DNA; 4559 BP.
 XX ABL18477;
 AC ABL18477;
 XX 26-MAR-2002 (first entry)
 DT Drosophila melanogaster genomic polynucleotide SEQ ID NO 6904.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX Drosophila melanogaster.
 OS Drosophila melanogaster.
 XX WO200171042-A2.
 XX 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR New isolated nucleic acid detection reagent for detecting 1000 or more

QY 1893 CTACGAGCCCTCTAGAAAACCTGACTGCAATTCCTGAGCATCGAATGACTACTACTACAAG 1952
DB 2785 CTGAAGAACATCGTTGAGAAAGTTGGCTGTGATAGCGGAGCACCGCATTTGATGTCTCATCAAG 2844
QY 1953 GCAAGTGAATAATTACATCTCTGTAGTAGATACACAGGTCACAGCTCAAAATTTCTTTGAAAAG 2012
DB 2845 TTGATCCACGCTATGAGCCCGCCACAGGATGTGCGGCTCAGATCAAGTTTCTCGAGAG 2904
QY 2013 CTGATCAATTTGGAGACGAGAGAAAGATTTGGAAGAAAGAGAAATTTTACTTAAGGCA 2072
DB 2905 CTGACAAAGCGCGAGCAGAGAGCAGAGGAACCTGGAGCTGAGATGCTGCTGCGGCA 2964
QY 2073 GCCAGAGTCTTCTTAATAAGAGATCCAGAACAGCTGAGATTAAGCAGAAAGCCAAA 2132
DB 2965 GCCAAGTCAAGCTGAGGTTGGAAGATCCCGAGCAGGCAAGATGAAGCGGAGGCGCAAG 3024
QY 2133 GAGTTACAGCAATTTGGAACTTGCACAGATACAGATAGAGACGCTAATCTCACAGCTCTT 2192
DB 3025 GAGATGCAACGCGCGGAATGGAGGAGTTGGCTCAACGAGATGCCATCTGACGCGCTG 3084
QY 2193 GCAGCTATTGGACCAAGGAAGAGACCACTAGAAATTTGAGGGCTTTAAAGAC 2252
DB 3085 CAGCGATTTGACCTCGGAAAAGCTGAAGCTGGAGCGGCAACAGTCAGTTCTGGAGCG 3144
QY 2253 AACCTTCTGCTTCTGG---GACATCCAGCTGACAGCCCAACAGTTGCATCGTCCA 2309
DB 3145 GGTTCAGTGGCGCGGAGTCTAAGCAGCTCGGGATCTGCGCGCAGCAGCTTACGGCCT 3204
QY 2310 AGAATCAGGAAATCTGCTCAGGAGCTTGATATTTTGTATGAACAGAGGAGGAGATG 2369
DB 3205 CGCATAAACGTGTGAACCTCGCGACATGCTCTTACATGGAGCAGAGCGGAGTTC 3264
QY 2370 AAGTATCTCGAGCTATACCTGGCCCTTCTGAAGTGACACACTCCACTCTTCCATCCA 2428
DB 3265 TGTGCGAGTTCCATGCTTCAAGACATACCTCAAGTATGCTGCTGCTGCCCATCAA 3323

RESULT 10

ID AAT42210 standard; cdna; 4615 BP.

XX AC AAT42210;

XX DT 27-JAN-1997 (first entry)

XX Drosophila TATA-binding protein associated factor dTAFII110 gene.

XX Drosophila; TATA-binding protein; TBP associated factor; TFIID;
KW RNA polymerase II; transcription; messenger RNA; nuclear fraction;
KW holoenzyme; lambda-gt11; expression library; ds.

XX OS Drosophila melanogaster.

XX FH Key Location/Qualifiers
XX CDS 538..3303
XX FT /*tag= a

XX FT /product= Drosophila TAFII110 protein

XX PN US5534410-A.

XX PD 09-JUL-1996.

XX PF 28-JAN-1993; 93US-0013412.

XX PR 28-JAN-1994; 94US-0188582.

XX PR 28-JAN-1993; 93US-0013412.

XX PR 30-JUN-1993; 93US-0087119.

XX PA (REGC) UNIV CALIFORNIA.

XX PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;

XX PI Tjian R, Wang E, Weinzierl ROJ;

XX DR

WPI: 1996-333245/33.

P-PSDB; AAW06077.

XX Screen for cpds. that bind human TATA-binding protein associated
PT factor - by testing ability to bind to polypeptide fragments of the
PT factor, useful as (ant)agonists of transcription factors involved in
PT disease.

XX Examples; Column 27-36; 86pp; English.

XX This is the nucleotide sequence encoding the Drosophila TATA-binding
CC protein (TBP) associated factor (TAF) designated TAFII110. The protein
CC is a component of the TFIID fraction required for reconstituting RNA
CC polymerase II in vitro transcription activity. The encoded protein
CC has an estimated mol. wt. of 110 kD by SDS-PAGE, a calculated mol. wt.
CC based on sequence of the gene of 99.4 kD and an estimated pI of 10.1.
CC The invention relates to purified proteins involved in transcription
CC by RNA polymerase II, the RNA polymerase which transcribes messenger
CC RNA. RNA polymerase II transcription proceeds in vitro upon addition
CC of several nuclear fractions designated TFIIA, B, D, E, F, H, I and J
CC to RNA polymerase II holoenzyme. Fraction TFIID has been shown to
CC contain a TBP and other TAFs. Purification of TFIID and separation of
CC its components reveals 7 proteins ranging in size from 30-250 kD.
CC Serum raised against the TFIID fraction allowed cloning of the corresp.
CC genes from lambda-gt11 expression libraries.

XX SQ Sequence 4615 BP; 1394 A; 1199 C; 1091 G; 931 T; 0 other;

Query Match 5.6%; Score 142.2; DB 17; Length 4615;

Best Local Similarity 50.7%; Pred. No. 4.1e-29;

Matches 395; Conservative 0; Mismatches 378; Indels 6; Gaps 2;

QY 1553 GATCAGGATGACATCAATGATGTGACTTCTATGGCAGGGGTCAACCTTTAATGAAGAAAT 1712
DB 2548 GGTGACGATGATATCAACGATGTTGCCGCCATGGGAGGTGTTAACTTGGCGGAGGATCG 2607

QY 1713 GCCTGTCATTTAGCAACAACTCTGAATGGTTGGCACACTCCTCAGTCACTCATGTAAGAT 1772
DB 2608 CAGCGAATTTCTCGGCTGTGTA---CCGAAACATCGCGCAGCATTCGATCTCTGCAAGAT 2664

QY 1773 GAACCATTTCTTTTATTGGAGCTCTACAAAAGAACTTTAGACATCTGTTAAAGAT 1832
DB 2665 GAGGTTTCTTAATCTCCCTCGCTGCAAGTAGAATACGGGCATTTACTTCGAGGCG 2724

QY 1833 GACATTACAGAACTTAATCTGATGCTGCTGAATCTCCCAAGCAACACAGGAACG 1892
DB 2725 GGACTGGATGAGCGCTCGCAGGATGTGCGCGTCTCTGATATCGCAGCGCTGTGAGGCGC 2784

QY 1893 CTACGAGGCTTCTAGAAAACTGACTGCATTTGCTCAGCATCGAATGACTACTTACAAG 1952
DB 2785 CTGAAGAACATCGTTGAGAAAGTTGGCTGTGATAGCGGAGCACCGCATTTGATGTCATCAAG 2844

QY 1953 GCAAGTGAATAATTACATCTGTGTAGTATACACAGTCCACAGCTCAAAATTTCTTGAAG 2012
DB 2845 TTGGATCCACGCTATGACCCCGCCAGGATGTGCGGCTCAGATCAAGTTTCTCGAGGAG 2904

QY 2013 CTGATCAATTTGGAGAACGAGAGAAAGATTTGGAAGAAAGAGAAATTTACTTAAAGCA 2072
DB 2905 CTGGACAAGCGCGAGCAGACGACAGGAACTGGAGCGTGTGATGCTGCTGCGGCA 2964

QY 2073 GCCAAGAGTCTTCTTAATAAGNAGATCCAGACAGCTGAGNTTAAGCAGAAAGCCAAA 2132
DB 2965 GCCAAGTCAAGCTCGAGGTTGGAAGATCCGAGCAGGCAAGATGAAGCGGAGGCGCAAG 3024

QY 2133 GAGTTACAGCAATTTGGAACCTTGACAGATACAGATAGAGAGCGCTAAATCTCACAGCTCTT 2192
DB 3025 GAGATGCAACGCGCGGAATGGAGGAGTTGCTCAACGAGATGCCAATCTGACGCGCTG 3084

QY 2193 GCAGCTATTGGACCAAGGAAGAGACCACTAGAAATCTGGAATTTGAGGGCTTTAAAGAC 2252
DB 3085 CAGCGATTTGACCTCGGAAAAGCTGAAGCTGGAGCGGCAACAGTCAGTTCTGGGAGCG 3144

QY 2253 AACCTTCTTCTTCTG---GACATCCAGCCTGACAGCCACCAACAGTTGATCGTCCA 2309
 Db 3145 GGTTCAGTGGCGGAGTGTAGCAGCTCGGATCTGCGCGACGACGTTACGGCT 3204
 QY 2310 AGAATCAGGAGATCTGCTCAGGGACTTGTATTTTGTATGGAACAGGAGGAGATG 2369
 Db 3205 CGCATAAACGTCGTAACCTGGCGACATGCTTCTTACATGAGCAAGCGGGAGTTC 3264
 QY 2370 AAGTATCTCGAGCTCTATACCTGGCCCTCTGAAGTGACCACTCCACATCTTCCATCCA 2428
 Db 3265 TGTCGAGTTCATGCTGTGTAAGACATACCTCAAGTGATGCTGCTGTGTCCTCA 3323

RESULT 11

AAT79604
 ID AAT79604 standard; cDNA; 4615 BP.

AC AAT79604;

DT 08-OCT-1997 (first entry)

DE TATA-binding protein associated factor, dTAFII110 cDNA.

KW TATA-binding protein associated factor; TAF; nuclear protein;

KW RNA polymerase transcription; TATA-binding protein; TBP;

KW initiation; ds.

XX Drosophila sp.

XX Key Location/Qualifiers

FT 538..3303

FT /*tag= a

XX US5637686-A.

XX 10-JUN-1997.

XX 28-JAN-1993; 93US-0013412.

XX 28-JAN-1994; 94US-0188582.

XX 28-JAN-1993; 93US-0013412.

XX 30-JUN-1993; 93US-0087119.

XX 09-MAY-1996; 96US-0646715.

XX (REGC) UNIV CALIFORNIA.

XX Comai L, Dynlact BD, Hoey T, Ruppert S, Tanese N;

PI Tjian R, Wang E, Weinzierl ROJ;

XX WPI; 1997-319113/29.

DR P-PSDB; AAW25028.

XX Nucleic acids encoding human TATA-binding protein associated factor

PT (TAF) peptide(s) - for production of recombinant peptide(s), used

PT for modulating transcription of TAFs

XX Example 1; Column 35-40; 86pp; English.

XX AAT79604 encodes Drosophila TATA-binding protein associated factor

CC (TAF) polypeptide, dTAFII110 (mol. weight 110kD). TAF peptides derived

CC from dTAFII30 alpha, dTAFII30 beta, dTAFII40, dTAFII60, dTAFII80,

CC dTAFII110, dTAFII150, and dTAFII250, their human equivalents and

CC nucleic acids encoding them, are used to modulate transcription,

CC including transcription initiation. TAFs are nuclear proteins involved

CC in RNA polymerase I, II and III transcription. The peptides act by

CC binding to a different TAF, an activator, or TBP (TATA-binding protein)

CC or competitively inhibiting association of a TAF domain with another

CC compound, typically a protein like TBP or another TAF, an activator,

CC or DNA.

XX Sequence 4615 BP; 1394 A; 1199 C; 1091 G; 931 T; 0 other;

XX Query Match 5.6%; Score 142.2; DB 18; Length 4615;

Best Local Similarity 50.7%; Pred. No. 4.1e-29;
 Matches 395; Conservative 0; Mismatches 378; Indels 6; Gaps 2;

QY 1653 GATGAGGATGACATCAATGATGTGACTTCTATGCGAGGGGTCAACCTTAATGAAGAAAT 1712
 Db 2548 GGTGAGCATGATATCAACGATGTTCGGCATGGGAGGTGTTAACTTGGCGGAGGATCG 2607
 QY 1713 GCTGTCATCTTAGCAACAACTCTGAATTTGGTGGCACACTCAATTCAGTCATCTAAAGAT 1772
 Db 2608 CAGCGAATCTCGGCTGTA---CCGAACAATCGGCACGAGATTCGATCTCTCAAGAT 2664
 QY 1773 GAACCATTTCTTTTATTTGAGCTCTACAAAAGAGAAATCTTAGACATTTGGTAAAAAGCAT 1832
 Db 2665 GAGGTTTTTCTTAATCTCCCTCGCTGCAAGCTAGAAATACGGCAATTAATCTCGGAGCG 2724
 QY 1833 GACATTACAGAACTTAATCTGATGCTGTGAATTTGATCTCCCAAGCAACACAGGACGA 1892
 Db 2725 GGACTGGATGAGCGGTCGCGGATGTGGCGGTCTCTATATCGCACGCCCTGTGAGGAGCG 2784
 QY 1893 CTACGAGGCTTTTAGAAAACTGACTGCAATTTGCTCAGCATCGAATGACTTACTTACAAG 1952
 Db 2785 CTGAAGAACATCGTTGAGAAGTTGGCTGTGATAGCGGAGCAGCCGATTTGATGTCATCAAG 2844
 QY 1953 GCAAGTGAATTTACATCTCTGTGTAGTGTATACAGGTCTACAGCTCAAAATTTCTTGAAG 2012
 Db 2845 TTGGATCCACGCTATGAGCCCGCCAAAGATGTGCGCGGTGATCAAGTAACTTCTCGAGGAG 2904
 QY 2013 CTGGATCAATTGAGAGAGCAGAGAAAGATTTGGAAGAAGAGAAATGTTACTTNAAGCA 2072
 Db 2905 CTGGACAGCGGACGAGAGAGAGACAGAGAACTGGAGCGTGATGATGCTGCTCGGGCA 2964
 QY 2073 GCCAAGTCTGTTCTTAATAAAGATCCAGAACACTGAGATTTAAAGCAGAAAGCCAAA 2132
 Db 2965 GCCAAGTCTGAGGTCGAGGTTGGAAGATCCGAGCAGCCCAAGATGAAGCGGAGGCCAAG 3024
 QY 2133 GAGTTACAGCAATTTGGAATTTGCACAGATACAGATAGACAGCTTAATCTCAGAGCTTT 2192
 Db 3025 GAGATGCAACCGCGCCGAAATGGAGGATTTGCGTCAACAGAGATGCCAATCTGAGCGGCTG 3084
 QY 2193 GCAGCTATTGGACCAAGAGAGAGACCACTAGATATCTGGAATTCAGGCTTAAAGAC 2252
 Db 3085 CAGCGGATTTGGACCTCGGAAAAGCTGGAAGCTGGAGCGGAAACAGCTCAGTTCGGAGCG 3144
 QY 2253 AACCTTCTTCTTCTG---GACATCCAGCCTGACAGCCACCAACAGTTGATCGTCCA 2309
 Db 3145 GGTTCAGTGGCGGAGTGTGCTAAGCAGCTCGGGATCTGCGCGACGACGTTACGGCT 3204
 QY 2310 AGAATCAGGAGATCTGCTCAGGGACTTGTATTTTGTATGGAACAGGAGGAGATG 2369
 Db 3205 CGCATAAACGTCGTAACCTGGCGACATGCTTCTTACATGAGCAAGCGGGAGTTC 3264
 QY 2370 AAGTATTTCTGAGCTCTATACCTGGCCCTCTGGAAGTGACCACTCCACATCTTCCATCCA 2428
 Db 3265 TGTCGAGTTCATGCTGTTCAGACATACCTCAAGTGATGCTGCTGTGTCCTCA 3323

RESULT 12

ABK32842

ID ABK32842 standard; DNA; 2307 BP.

XX AC ABK32842;

XX DT 23-APR-2002 (first entry)

XX DE DNA encoding human homologue of MPT1 antifungal target.

XX antifungal; fungal gene transcription; RPC34; POP3; TPA2; NAB2;

KW MPT1; MTR2; BOS1; POL30; RSA2; SQP1; MTW1; TFB1; SPC98; BFR2; RNA1;

KW GCD7; SKI6; NIP1; LCP5; NCE103; ECO1; ORC2; CNS1; YPD1; TIM10; SRB4;

XX yeast; fungus; ds; gene.

XX Homo sapiens.

OS

XX

PN WO200202055-A2.
 XX 10-JAN-2002.
 PD
 XX 28-JUN-2001; 2001WO-US20592.
 XX 29-JUN-2000; 2000US-215164P.
 XX 10-AUG-2000; 2000US-224457P.
 PR
 XX (ANAD-) ANADYS PHARM INC.
 PI Moore J, Buurman ET, Desilva T, Harris S, Komarnitsky S;
 PI Mendillo M, Moore D, McCoey M, Sanderson K, Haq T, Zhu S, Long F,
 PI Davidov E, Thompson CM;
 XX
 DR WPI; 2002-147962/19.
 DR P-PSDB; AAU82954.
 XX
 XX Screening candidate antifungal compound for interaction with essential
 PT protein, modulation of essential protein activity, binding to essential
 PT protein, by contacting protein with test compound and determining
 PT effects
 XX
 PS Disclosure; Figure 80; 522pp; English.
 XX
 CC The invention describes a method of screening a candidate antifungal
 CC compound for interaction with essential proteins (EP) or for modulation
 CC of EP activity e.g fungal gene transcription. The proteins tested in the
 CC invention include RPC34, POP3, TFA2, NAB2, MPT1, MTR2, BOS1, POL30, RSA2,
 CC SOT1, MTW1, TFB1, SPC38, BFR3, RNAL, GCD7, SKI6, NIP1, LCPS, NCE103,
 CC EGO1, ORC2, CNS1, YPD1, TIM10 and SRB4 from *S. cerevisiae*, *C. albicans*
 CC and human homologues. The method involves contacting a culture with one
 CC or more test compounds and determining the effects on the growth or
 CC viability of the culture of cells which preferably comprises fungal cells
 CC or yeast cells. Preferably the identified compounds interact with, or
 CC modulate (preferably inhibit) activity of *C. albicans* EP. The inhibitor
 CC compounds identified by the method are useful for preventing or
 CC inhibiting fungal, particularly *C. albicans* growth in culture or in a
 CC mammal. The antifungal agents interact with essential fungal elements
 CC that can be used to treat fungal infection by preventing the growth and
 CC preferentially killing the fungi, but does not inhibit the biological
 CC activity of mammalian homologues. This sequence encodes a target protein
 CC used to test the antifungal compounds, described in the method of the
 CC invention.
 XX
 SQ Sequence 2307 BP; 359 A; 969 C; 724 G; 255 T; 0 other;
 Query Match 5.3%; Score 135.8; DB 24; Length 2307;
 Best Local Similarity 66.7%; Pred. No. 1.8e-27;
 Matches 194; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
 QY 556 TTCTCCGACAAATGCTAGAAATGTGAAGAAATGCAAGAACTTCCTTCAATGTTAATAAA 615
 DB 1776 TGGCAGGAAACTATGGAACACGTGAAGAAATGTAAAAATTTCTATCTAGCTTAATAAA 1835
 QY 616 ACTAGCATGTAGTGGATCACAGTCCCTGAAATGGGCAAAATGTGAAGAGCTGGTGA 675
 DB 1836 ACTGGCTTCATCTGGCAAGCACTCTACAGACAGACAGCTAATGTGAAGAGCTCGTGCA 1895
 QY 676 ACAACTTTTGGATCAAAAATCGAAGACAGAGAAATTTACTAGGAAACTGTATGTTGAAC 735
 DB 1896 GAATTTACTGGATGAAATAATGAGACAGAGATTTTCAACAGCAGGTATATACCGAGAACT 1955
 QY 736 CAAGCTTTTCACTCCAGCTCACCTGGTTCCTTTTCAAGAAAACGCTGGTTCCTTACG 795
 DB 1956 TAATTTCTTCACTCACTCACTTGTGCTTTCCTTGAAGAGAGCTTACCCGCTTGAG 2015
 QY 796 ACAACTTCTGCCTTAACCTCCAGAGCTTCATCCACCAATGTGTTTCAGCAGAC 846
 DB 2016 ACAGCTGACCCCGACACTCCGCGCTTTCATCCAGCAGAGCAGCAGCAGC 2066

RESULT 13

ABN33770
 ID ABN33770 standard; DNA; 60 BP.
 XX
 AC ABN33770;
 XX
 DT 15-JUL-2002 (first entry)
 XX
 DE Human spliced transcript detection oligonucleotide SEQ ID NO:6518.
 XX
 KW Human; mouse; rat; splice transcript; detection; RNA transcript;
 KW splice variant; transcripome; oligonucleotide library; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200210449-A2.
 XX
 PD 07-FEB-2002.
 XX
 PF 20-JUL-2001; 2001WO-IB01903.
 XX
 PR 28-JUL-2000; 2000US-221607P.
 PR 02-MAY-2001; 2001US-287724P.
 XX
 PA (COMP-) COMPUGEN INC.
 XX
 PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
 XX
 DR WPI; 2002-257383/30.
 XX
 PT New oligonucleotide libraries comprising oligonucleotides which
 PT selectively hybridize to mRNAs transcribed from a transcription unit of
 PT a genome, useful for detecting tissue-, pathology-, and
 PT developmental-specific genes
 XX
 PS Example 1; SEQ ID 6518; 47pp; English.
 XX
 CC The present invention describes oligonucleotide libraries for detecting
 CC messenger RNAs that populate a (sub-)transcriptome, where the
 CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
 CC transcription units that populate a genome. The library comprises
 CC several oligonucleotides, each capable of hybridizing selectively to a
 CC set of messenger RNAs transcribed from a given transcription unit of
 CC the genome, which encodes one or more messenger RNA splice variants.
 CC The oligonucleotide libraries are useful for detecting mRNAs from a
 CC biological sample, in expression profiling studies, in qualitatively or
 CC quantitatively characterising the corresponding transcriptome, and in
 CC detecting RNA transcripts and splice variants of human or animal
 CC transcriptomes. The libraries may also be used as specialised mini
 CC libraries to detect transcripts of a sub-transcriptome under a
 CC particular biological or pathological state, and so allowing the
 CC detection of tissue- and pathology-specific genes such as those genes
 CC only expressed in specific tissue under a specific pathological
 CC condition; to detect developmental specific genes; and to detect RNA
 CC transcripts and splice variants of a transcriptome of a patient suffering
 CC from a particular disorder. ABN27253 to ABN59589 represent
 CC oligonucleotide sequences from rats, humans and mice, which are used in
 CC the exemplification of the present invention.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 60 BP; 17 A; 17 C; 11 G; 15 T; 0 other;
 Query Match 2.3%; Score 60; DB 24; Length 60;
 Best Local Similarity 100.0%; Pred. No. 6.6e-07;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 873 GTACTACACAGTAACTTCTCTCTGTGGTACAACTACAGTGTCTCAAGCCAGTCTG 932
 DB 1 GTACTACACAGTAACTTCTCTCTGTGGTACAACTACAGTGTCTCTCAAGCCAGTCTG 900

RESULT 14

ABLO9422/c
ID ABL09422 standard; cDNA; 9525 BP.
XX AC ABL09422;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 22748.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WIPI; 2001-656860/75.
XX DR P-PSDB; ABB65319.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Claim 1; SEQ ID NO 22748; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLO1840-ABL16175), expressed DNA
CC sequences (ABLO1840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 9525 BP; 2503 A; 1867 C; 2248 G; 2907 T; 0 other;
Query Match 2.3%; Score 58.6; DB 23; Length 9525;
Best Local Similarity 51.4%; Pred. No. 2.5e-05;
Matches 161; Conservative 0; Mismatches 149; Indels 3; Gaps 1;
QY 1653 GATGAGGATGACATCAATGATCTGACTTCTATGCGAGGGTCAACCTTAATGAAGAAAT 1712
DB 3243 GGTGAGGATGATCAACGATGTGCCGCCATGGGAGGTGTTAACTTGGCGGAGGATCG 3184
QY 1713 GCCTGCATCTTAGCAACAACCTCTGAATTGGTTGGCACACTCATTCAGTCATGTAAGAT 1772
DB 3183 CAGCGAATTTCTCGGCTGTA---CCGAAACATCGGCAGCATTCGATTCCTGCAAGAT 3127
QY 1773 GAACCATTTCTTTTATGGAGCTCTACAAAAGAAATCTTAGACATTTGGTAAAGCAT 1832
DB 3126 GAGGATTTCTCAATCTCCCTCGCTGCAAGCTAGAACTAGCGGCAATTAATCTCGGAGGCG 3067
QY 1833 GACATTACAGAACTTAACCTCTGATGCTGTGAATTCCTCCAGCAACACAGCAACGA 1892
DB 3066 GACATGAGATGAGCCGTGCGAGGATGTGCCGCTTCTGATATCGACGCGCTGTGAGGAGCG 3007
QY 1893 CTACGAGGCGCTTCTAGAAAACTGACTGCAATTTGCTCAGCATCGAATGACTACTTACAG 1952
DB 3006 CTCAGAACATCGTTGAGAACTTGGCTGTGATAGCGGAGCACCAGCATTCATCATCAG 2947

QY 1953 GCAAGTCAAAAT 1965
DB 2946 GTTTGTGGAAAT 2934
RESULT 15
ABL18476/c
ID ABL18476 standard; DNA; 10083 BP.
XX AC ABL18476;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 6901.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WIPI; 2001-656860/75.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Claim 1; SEQ ID NO 6901; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 10083 BP; 2703 A; 1962 C; 2336 G; 3082 T; 0 other;
Query Match 2.3%; Score 58.6; DB 23; Length 10083;
Best Local Similarity 51.4%; Pred. No. 2.5e-05;
Matches 161; Conservative 0; Mismatches 149; Indels 3; Gaps 1;
QY 1653 GATGAGGATGACATCAATGATGTGACTTCTATGCGAGGGTCAACCTTAATGAAGAAAT 1712
DB 3243 GGTGAGGATGATCAACGATGTGCCGCCATGGGAGGTGTTAACTTGGCGGAGGATCG 3184
QY 1713 GCCTGCATCTTAGCAACAACCTCTGAATTGGTTGGCACACTCATTCAGTCATGTAAGAT 1772
DB 3183 CAGCGAATTTCTCGGCTGTA---CCGAAACATCGGCAGCATTCGATTCCTGCAAGAT 3127
QY 1773 GAACCATTTCTTTTATGGAGCTCTACAAAAGAAATCTTAGACATTTGGTAAAGCAT 1832
DB 3126 GAGGATTTCTCAATCTCCCTCGCTGCAAGCTAGAACTAGCGGCAATTAATCTCGGAGGCG 3067
QY 1833 GACATTACAGAACTTAACCTCTGATGCTGTGAATTCCTCCAGCAACACAGCAACGA 1892

```

Db 3066 G G A C T G G A T G A G C C G T C G C A G G A T G T G G C C G T T C T G A T A T C G C A G C C C T G T C A G G A G G C 3007
Qy 1893 C T A C G A G C C C T T C T A G A A A A A C T G A C T G C A A T T G C T A G C A T C G A A T G C A T G A C T A C T T A C A A G 1952
Db 3006 C T G A A G A A C A C T C G T T G A G A A G T T G G C T G T G A T A G C G G A G C A C C G C A T T G A T G T C A T C A A G 2947
Qy 1953 G C A A G T G A A A A T T 1965
Db 2946 G T T T G T T G G A A A T 2934

```

Search completed: February 16, 2003, 22:28:41
 Job time : 631 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:54:50 ; Search time 16.4197 Seconds
(without alignments)
2152.161 Million cell updates/sec

Title: US-09-763-909-2

Perfect score: 4264

Sequence: 1 GTLVTKVAPVSAPPKVSSCP.....KZQHERALFTIRTLTLTY 852

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3990	93.6	801	1 T2DT_HUMAN	Q92750 homo sapien
2	1307.5	30.7	1083	1 T2D3_HUMAN	P00268 homo sapien
3	775	18.2	921	1 T2D3_DROME	P47825 drosophila
4	239.5	5.6	2035	1 HFCL_HUMAN	P51610 homo sapien
5	238	5.6	2090	1 HFCL_MESAU	P51611 mesocricetu
6	220	5.2	1367	1 AMYH_YEAST	P08640 saccharomyc
7	219	5.1	5179	1 MUC2_HUMAN	Q02817 homo sapien
8	213.5	5.0	5376	1 ZAN_MOUSE	O88799 mus musculus
9	206.5	4.8	865	1 CPN_DROME	Q02910 drosophila
10	205.5	4.8	1161	1 DAN4_YEAST	P47779 saccharomyc
11	203.5	4.8	1322	1 YAG3_YEAST	P39712 saccharomyc
12	201	4.7	725	1 AGAL_YEAST	P32323 saccharomyc
13	199	4.7	2700	1 ZAN_HUMAN	Q9Y493 homo sapien
14	197.5	4.6	881	1 YJH8_YEAST	P47033 saccharomyc
15	197.5	4.6	1609	1 FIG2_YEAST	P25653 saccharomyc
16	193.5	4.5	670	1 VG50_HSV1	Q00130 ictaluriid h
17	191.5	4.5	1260	1 ALS1_CANAL	P46590 candida alb
18	191	4.5	797	1 VGLX_HSVB	P28968 equine herp
19	187.5	4.4	3726	1 TRX_DROME	P20659 drosophila
20	186	4.4	3178	1 YS89_CABEL	Q09624 caenorhabdi
21	184.5	4.3	1537	1 FLO1_YEAST	P32768 saccharomyc
22	182	4.3	1075	1 FLO5_YEAST	P38894 saccharomyc
23	182	4.3	3866	1 HRX_MOUSE	P55200 mus musculus
24	181.5	4.3	662	1 MUC1_XENLA	Q05048 xenopus lae
25	181.5	4.3	1199	1 N121_RAT	P52591 rattus norv
26	181	4.2	1858	1 P3K2_DICDI	P54674 dictyostelli
27	177.5	4.2	1140	1 YM96_YEAST	Q04893 saccharomyc
28	176	4.1	2541	1 TALI_HUMAN	Q9Y490 homo sapien
29	175.5	4.1	2090	1 N214_HUMAN	P33658 homo sapien
30	175.5	4.1	3969	1 HRX_HUMAN	Q03164 homo sapien
31	173.5	4.1	1802	1 HKR1_YEAST	P41809 saccharomyc
32	173	4.1	2404	1 SON_MOUSE	Q9qx47 mus musculus
33	172	4.0	886	1 VGP3_EBVA8	Q07284 epstein-bar

34	172	4.0	1018	1 HMWL_MYCPN	O50365 mycoplasma
35	172	4.0	1119	1 ALS3_CANAL	O74623 candida alb
36	172	4.0	1169	1 YK82_YEAST	P36170 saccharomyc
37	171.5	4.0	1780	1 YK26_CABEL	P34333 caenorhabdi
38	170	4.0	1306	1 MSB2_YEAST	P32334 saccharomyc
39	169.5	4.0	388	1 MPT1_YEAST	P50105 saccharomyc
40	169.5	4.0	745	1 OCT1_PIG	O29076 sus scrofa
41	169.5	4.0	2426	1 SON_HUMAN	P18583 homo sapien
42	169	4.0	1229	1 N121_HUMAN	Q9Y2H3 homo sapien
43	166.5	3.9	743	1 OCT1_HUMAN	P14859 homo sapien
44	165	3.9	606	1 SP2_HUMAN	Q02086 homo sapien
45	165	3.9	2541	1 TALI_MOUSE	P26039 mus musculus

ALIGNMENTS

RESULT 1

ID	T2DT_HUMAN	STANDARD;	PRT;	801 AA.
AC	Q92750;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Transcription initiation factor TFIIID 105 kDa subunit (TAFII-105)			
DE	(TAFII105) (Fragment).			
GN	TAF4B OR TAF2C2 OR TAFII105.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE:97011146; PubMed:8858156;			
RA	Dikstein R., Zhou S., Tjian R.;			
RT	"Human TAFII 105 is a cell type-specific TFIIID subunit related to hTAFII130."			
RL	Cell 87:137-146(1996).			
CC	-1- FUNCTION: CELL TYPE-SPECIFIC SUBUNIT OF TFIIID THAT MAY FUNCTION AS A GENE-SELECTIVE COACTIVATOR IN CERTAIN CELLS. TFIIID IS A MULTIMERIC PROTEIN COMPLEX THAT PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS ACTIVATORS AND REPRESSORS.			
CC	-1- SUBUNIT: TFIIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A NUMBER OF TBP-ASSOCIATED FACTORS (TAFs). TAFII105 MAY FORM HETERODIMERS WITH TAFII130.			
CC	-1- SUBCELLULAR LOCATION: Nuclear.			
CC	-1- TISSUE SPECIFICITY: FOUND IN B CELLS BUT NOT IN OTHER TESTED CELLS.			
CC	-1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR	EMBL; Y09321; CAA70499.1; -
DR	Genew; HGNC:11538; TAF4B.
DR	MTM; 601689; -
DR	InterPro; IPR003894; TAF_hom.
DR	SMART; SM00549; TAFH; 1.
KW	Transcription regulation; Nuclear protein.
FT	NON_TER 1
FT	NON_TER 1
SQ	SEQUENCE 801 AA; 85658 MW; D12B4932FEA9CD2 CRC64;

Query Match 93.6%; Score 3990; DB 1; Length 801;
Best Local Similarity 99.9%; Pred. No. 5.1e-199;
Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTLVTKVAPVSAPPKVSSGRLPAPQIVAVKAPNTTITQIPANQLPLPGTVLIKNSGCL 60
|||||

Db 1 GTLVTKVAPVSAPPKVSQGRPLPAPQIVAVKAPNTTIIQFPANLQIPGCTVLKNSGSL 60
 QY 61 MLVSPQOTVTRAEFTSNITSRPAVPANPOTVKTCTVPNSSSOLIKKVAVTPVKKLAQIGT 120
 Db 61 MLVSPQOTVTRAEFTSNITSRPAVPANPOTVKTCTVPNSSSOLIKKVAVTPVKKLAQIGT 120
 QY 121 TVVTVTPKSSVQSVAVPTSVTVTPGKPLNTVTTLKPSLIGASSTPSNEPLKKAENSA 180
 Db 121 TVVTVTPKSSVQSVAVPTSVTVTPGKPLNTVTTLKPSLIGASSTPSNEPLKKAENSA 180
 QY 181 VOINLSPMLENVKCKNFELAMLIKACSGSPENGVNKKVLQGLDIAKEAEFTTRK 240
 Db 181 VOINLSPMLENVKCKNFELAMLIKACSGSPENGVNKKVLQGLDIAKEAEFTTRK 240
 QY 241 LVEKSSQPHLVPFLPKSVKVALROLPLNSQSFIOQCQVQSTSDMWIATCTTTTSPV 300
 Db 241 LVEKSSQPHLVPFLPKSVKVALROLPLNSQSFIOQCQVQSTSDMWIATCTTTTSPV 300
 QY 301 VTTTVSSSSEKSIIVSGATARTVSQTLNPLAGPVGAKAGVTVLHSHVGPPTAATGGTTA 360
 Db 301 VTTTVSSSSEKSIIVSGATARTVSQTLNPLAGPVGAKAGVTVLHSHVGPPTAATGGTTA 360
 QY 361 GTGLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAVTSLPVTFTGETSGAAICLPSVKP 420
 Db 361 GTGLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAVTSLPVTFTGETSGAAICLPSVKP 420
 QY 421 VVSFCDWHICKPVIPTVQVQIKLQAGPVLSPAGIPTGSSSKQLFSLFHVVOQPSGGNEK 480
 Db 421 VVSFCDWHICKPVIPTVQVQIKLQAGPVLSPAGIPTGSSSKQLFSLFHVVOQPSGGNEK 480
 QY 481 QVTTTSHSSTLTIOCKGOKTQVNTIIPTSQPPASILKQITLPGNKILSLQASPTQNR 540
 Db 481 QVTTTSHSSTLTIOCKGOKTQVNTIIPTSQPPASILKQITLPGNKILSLQASPTQNR 540
 QY 541 IRENTVTSRDEDDINDVTSAGVNLNEENACILATNSLVTGLTQSKDEFFLFTGALQ 600
 Db 541 IRENTVTSRDEDDINDVTSAGVNLNEENACILATNSLVTGLTQSKDEFFLFTGALQ 600
 QY 601 KRILQIKKHDTIENSDAVNLISQATQRLRGLEKLTATIAQHRMTYTKASENVLCS 660
 Db 601 KRILQIKKHDTIENSDAVNLISQATQRLRGLEKLTATIAQHRMTYTKASENVLCS 660
 QY 721 QHRDANLTAAIGPKRKRPLESGIEGLKNDLLASGTSLSLTATKQHRPRITRICLDLI 780
 Db 721 QHRDANLTAAIGPKRKRPLESGIEGLKNDLLASGTSLSLTATKQHRPRITRICLDLI 780
 QY 781 FCMEQEREMKYSRALLYALLK 801
 Db 781 FCMEQEREMKYSRALLYALLK 801

RESULT 2

T2D3_HUMAN
 ID T2D3_HUMAN STANDARD; PRT: 1083 AA.
 AC 000268; Q99721; Q9BX42; Q9BR40;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcription initiation factor TFIIID 135 kDa subunit (TAFII-135)
 DE (TAFII135) (TAFII-130) (TAFII130).
 GN TAF4 OR TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP MEDLINE=97336072; PubMed=9192867;
 RX Mengus G., May M., Carre L., Chambon P., Davidson I.;

RT

RT

RT

RL

RN

RN

RP

RX

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

RA

Human TAF(II)135 potentiates transcriptional activation by the AP-2s of the retinoic acid, vitamin D3, and thyroid hormone receptors in mammalian cells.;
 Genes Dev. 11:1381-1395(1997).

SEQUENCE FROM N.A.
 MEDLINE=21638749; PubMed=11780052;
 DeLoukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R., Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagdley C.L., Bailey J., Barlow K.F., Bates K.N., Beare D.M., Beare D.M., Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J., Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M., Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R., Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M., Ellington A.G., Frankland J.A., Fraser A., French L., Garner P., Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E., Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J., Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D., Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S., Leharshaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D., Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A., Milne S.A., Misty D., Moore M.J.F., Mullikin J.C., Nickerson T., Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I., Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H., Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S., Skuce C.D., Smith M.H., Soderlund C., Steward C.A., Sulston J.E., Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D., Thorpe A., Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A., Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S., Rogers J.;
 The DNA sequence and comparative analysis of human chromosome 20.;
 Nature 414:865-871(2001).
 [3]
 SEQUENCE OF 105-1083 FROM N.A. AND PARTIAL SEQUENCE.
 MEDLINE=97098442; PubMed=8942982;
 Tanese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;
 Molecular cloning and analysis of two subunits of the human TFIIID complex: hTAFII130 and hTAFII100.;
 Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).
 CC -1- FUNCTION: MAKES PART OF TFIIID IS A MULTIMERIC PROTEIN COMPLEX THAT
 CC ACTIVATORS AND REPRESSORS. POTENTIATES PROMOTER RESPONSES TO VARIOUS
 CC BY THE AP-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.
 CC -1- SUBUNIT: TFIIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
 CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; Y11354; CAA72189.1; -

EMBL; AL137077; CAC36006.1; -

EMBL; AL109911; CAC23212.2; -

EMBL; U75308; AAC50901.1; -

TRANSFAC; T02328; -

Gene; HGNC:11537; TAF4.

MIM; 601796; -

InterPro; IPR003894; TAF_hom.

SMART; SM00549; TAFH; 1

Transcription regulation: Nuclear protein.

DOMAIN 39 42 POLY-HIS.

DOMAIN 52 57 POLY-ALA.

DOMAIN 98 101 POLY-GLY.

DOMAIN 142 148 POLY-ALA.

DOMAIN 268 275 POLY-PRO.

RESULT 3

TD3_DROME STANDARD; PRT; 921 AA.

ID T2D3_DROME STANDARD; PRT; 921 AA.

AC P47825; P49845; Q9VUV7;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

DE Transcription initiation factor TFIID 110 kDa subunit (P110)

DE (TAFII-110) (110 kDa TBP-associated factor).

GN TAF110 OR G5444.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RI [1]

RN TAF110 OR G5444.

RP Drosophila melanogaster (Fruit fly).

RC TISSUE=Embryo;

RC TISSUE=Embryo;

RX MEDLINE=93145326; PubMed=7678780;

RY Hoey T., Weinzierl R.O.J., Gill G., Chen J.-L., Dynlacht B.D.,

RA Tjian R.;

RT "Molecular cloning and functional analysis of Drosophila TAF110

RT reveal properties expected of coactivators.";

RL Cell 72:247-260(1993).

RI [2]

RN TAF110 OR G5444.

RP Drosophila melanogaster (Fruit fly).

RC TISSUE=Embryo;

RC TISSUE=Embryo;

RX MEDLINE=93317591; PubMed=8327460;

RY Kokubo T., Gong D.-W., Roeder R.G., Horikoshi M., Nakatani Y.;

RA "The Drosophila 110-kDa transcription factor TFIID subunit directly

RA interacts with the N-terminal region of the 230-kDa subunit.";

RL Proc. Natl. Acad. Sci. U.S.A. 90:5896-5900(1993).

RI [3]

RN TAF110 OR G5444.

RP Drosophila melanogaster (Fruit fly).

RC TISSUE=Embryo;

RC TISSUE=Embryo;

RX MEDLINE=20196006; PubMed=10731132;

RY Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Ananides P.G., Scher S.E., Richards S., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Edwards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers J.-H.C., Blake J.G., Champagne M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Baller R.M., Basu A.A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,

RA Finkler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glöckner A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Moadar C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J.J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F., Zhong W., Zhou X., Zhu S., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster."

RL Science 287:2185-2195(2000).
 CC -!- FUNCTION: TFIIID IS A MULTIMERIC PROTEIN COMPLEX THAT PLAYS A
 CC CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS ACTIVATORS
 CC AND REPRESSORS. MAY FUNCTION AS A COACTIVATOR BY SERVING AS A SITE
 CC OF PROTEIN-PROTEIN CONTACT BETWEEN ACTIVATORS LIKE SPI (OR BTD)
 CC AND TFIIID COMPLEX.
 CC -!- SUBUNIT: TFIIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
 CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; J06861; -; NOT_ANNOTATED_CDS.
 DR EMBL; S63550; AAB27433.1; -.
 DR EMBL; AB003526; AAF49536.1; -.
 DR TRANSFAC; T02121; -.
 DR FlyBase; FBgn0010280; Taf110.
 DR InterPro; IPR003894; TAF_hom.
 DR SMART; SM00549; TAFH; 1.
 KW Transcription regulation; Nuclear protein.
 FT DOMAIN 66 82 POLY-GLN.
 FT DOMAIN 108 111 POLY-GLN.
 FT DOMAIN 259 265 POLY-GLN.
 SQ SEQUENCE 921 AA; 93338 MW; 27E6852899872767 CRC64;
 Query Match 18.2%; Score 775; DB 1; Length 921;
 Best Local Similarity 29.6%; Pred. No. 7.2e-33;
 Matches 258; Conservative 128; Mismatches 308; Indels 178; Gaps 26;
 QY 25 PQIVAVKAEPTTIQPPANLQPPGTVLKSNGPML-----VSPQQVTRAEYTS 76
 DB 133 POSPSITLSTLNTGQTPA-----LLVKTONGFQLLRVGTGTPPTVTQITNTSNN 184
 QY 77 NITSRAVPANPQTVKICVTPNSSSQ-----LIKKVAVTPVKKLAQIGTVVTVTP 127
 DB 185 NITSNTNHTTQ-Q-ILQTVPAASMTNTATNTATNIIVNSVASSGVNSQPHLQQLNAQ 243
 QY 128 KPSSVQSVAVPTSVVTVTPGKPLNTVTLKPSLSGLSSTPSNEPNLKAENSAVQINLSP 187
 DB 244 APQLPQITQITIPAOQSOQOQVNNVSSAGGTATAVSSTA-----ATT 287
 QY 188 TMLNVK-KCKNFLAMLIKAGSQSPENGVKVLVEQLLDKAEAEFTKLYVELK 246
 DB 288 TQGNTEKCKRFLANLIEL--STREPKPVEKNVRTLQELVNAVPEPEFCDRLERLEN 345
 QY 247 SSPQHLVPLFKSVVALROL-----LPNSQSFTQ--- 276
 DB 346 ASPQCLIGFLKSLPLLRQALYTKELVIEGKPPQHVGLAGLSQQLPKIQAIRPG 405
 QY 277 ----OCVQOTSDMWIATCTTTTTPVTVTVSSSQSEKSIIVSGATPRTVS---VQT 329
 DB 406 PSQTTTIGTQVRMI--TPNALCTPRPTTIGHTTISKQPNP---IRLPTAPRLVNTGGIRT 460
 QY 330 LNPAGPVCAGAGVTLHSVGTATAGTTAGTGLQTSKPLVTSVANTVTVTSVLOPEKP 389
 DB 461 QIP-SLQVPGQANVOIR--GPOHAQLQRTGSGVQIRATRP-----PNSVPTAN----- 506
 QY 390 VVSGTAVTSLPAVTFGETSGAICLPSVKPVVSCWDHICKPVIGTPVQIKLAOPGVL 449
 DB 507 -----KLRAVVGQTIKAI-TPSLHP-----PSLAISGGP-----PPTPL 543
 QY 450 SQAPAGTSSSK-----QLFSLFVVVQVQSPGGNEKQVTTISHS----- 489
 DB 544 SVLSTLNSASTTTLPIPLPTVHLPEALRAREQMSLNHNHNFDAKLVEIKAPSLHP 603
 QY 490 -----TLTIQCGKQKTPVNTIITPSPPPAKILKQITLPGNKILSQ-----AS 534

RESULT 4

HFCL_HUMAN
 ID HFCL_HUMAN STANDARD; PRT; 2035 AA.
 AC P51610.
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Host cell factor C1 (HCF) (VP16 accessory protein) (HFCL) (VCAF)
 DE (CFF).
 GN (HCF1 OR HCF1)
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS), AND PARTIAL SEQUENCE.
 RX MEDLINE=9327419; PubMed=8392914;
 RA Wilson A.C., Lamarco K., Peterson M.G., Herr W.;
 RT "The VP16 accessory protein HCF is a family of polypeptides processed
 RT from a large precursor protein.";
 RL Cell 74:115-125(1993).
 RN [2]
 RP SEQUENCE OF 65-2035 FROM N.A. (LONG AND SHORT ISOFORMS).
 RC TISSUE=Fetal brain;
 RX MEDLINE=95130107; PubMed=7829097;
 RA Frattini A., Faranda S., Redolfi E., Zucchi I., Villa A.,
 RA Patrosso M.C., Strina D., Susani L., Vezzoni P.;
 RT "Genomic organization of the human VP16 accessory protein, a
 RT housekeeping gene (HCF1) mapping to Xq28.";
 RL Genomics 23:30-35(1994).
 RN [3]
 RP PROCESSING.
 RX MEDLINE=96033796; PubMed=7590226;
 RA Wilson A.C., Peterson M.G., Herr W.;
 RT "The HCF repeat is an unusual proteolytic cleavage signal.";
 RL Genes Dev. 9:2445-2458(1995).
 CC -!- FUNCTION: UPON LYTC INFECTION OF PERMISSIVE CELLS, THE HSV
 CC TRANSACTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF
 CC ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING
 CC PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR
 CC ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES.
 CC -!- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM
 CC 110 TO 150 kDa AND A MINOR 300 kDa POLYPEPTIDE. THE MAJORITY OF N-
 CC AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBEIT
 CC NONCOVALENTLY, ASSOCIATED.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORTER FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING. IN THE
 CC SHORTER FORM THE N- AND THE C-TERMINAL FRAGMENTS FAIL TO

CC ASSOCIATE.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PETAL TISSUES AND THE
CC ADULT KIDNEY. PRESENT IN ALL TISSUES TESTED.
CC -1- DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE
CC SIGNAL.
CC -1- PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH
CC CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS
CC AT A DEFINED SITE, PPCE/THET, WITHIN THE HCF REPEAT.
CC -1- PTM: GLYCOSYLATED; CONTAINS O-LINKED N-ACETYLGLUCOSAMINE RESIDUES.
CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L20010; ; NOT_ANNOTATED_CDS.
CC EMBL: X79198; CAA55790.1; AUT_INIT.
CC Genew: HGNC:4839; HCFCL.
CC MIM: 300019; .
CC InterPro: IPR003961; FN.III.
CC InterPro: IPR001798; Kelch.
CC Pfam: PF00041; fn3; 1.
CC Pfam: PF01344; Kelch; 5.
CC SMART: SM00060; FN3; 1.
CC Nuclear protein; Repeat; Alternative splicing; Glycoprotein.
CC
CC FT REPEAT 44 69
CC FT REPEAT 93 140
CC FT REPEAT 148 194
CC FT REPEAT 217 265
CC FT REPEAT 266 313
CC FT DOMAIN 1010 1439
CC FT REPEAT 1010 1035
CC FT REPEAT 1072 1097
CC FT REPEAT 1101 1126
CC FT REPEAT 1158 1183
CC FT REPEAT 1286 1311
CC FT REPEAT 1314 1339
CC FT REPEAT 1349 1374
CC FT REPEAT 1414 1439
CC FT VARSPLIC 382 450
CC FT CONFLICT 564 564
CC FT CONFLICT 603 603
CC FT CONFLICT 1164 1164
CC FT CONFLICT 1873 1873
CC FT SEQUENCE 2035 AA; 208841 MW; 99207FBE875204C0 CRC64;
CC
CC Query Match 5.6%; Score 239.5; DB 1; Length 2035;
CC Best Local Similarity 22.6%; Pred. No. 9.4e-05;
CC Matches 164; Conservative 89; Mismatches 245; Indels 229; Gaps 38;
CC
CC QY 2 TLTKVAPVSAPPKVSRRPLP-----APQIVAVKAPNTT-----IQ 39
CC
CC Db 446 TLTPQAAP--APTPTTIIQVLPVPGSSISVPTAARTQGVPAVLKVTGPQATTGTPLVTM 503
CC
CC QY 40 FPNAN-----LQLPPGTVLI---KNSGGLMLVSPQOT-----VTRAETTSNI--TS 80
CC
CC Db 504 RPASQAGKAPVTYSLPAGVVMVPTQSAQGTIVGSSPQMSGMAALAAAAATQKTPSS 563
CC
CC QY 81 RP---AVPAN-----PQTVKICTVP-----NSSQLIKKVVAVTPVKKLAQ 117
CC
CC Db 564 RPTVLSVPAGTTIVKTMVTPGTTLTPATVKVASSPVMVSNPATRMLKATA-----AQ 616
CC
CC QY 118 IGTTV-----VTVPKPSVSVQSVAVPTSVTVTPKPLNTVTLK-PSGL-CASS 165
CC
CC Db 617 VGVSVSSATNTSRPTIIVHKSQTV-TVAQQAQVTVTVVGVGVTKTITLVKSPISVPGGSA 675
CC
CC QY 166 TPNENPL-----KAENSAAVQINLSPTMLENVKCKNPL--AMLIKACSGSQSP--- 214
CC
CC Db 676 LISNLKVMVSVQTKPVQISAVTQASSTGPTVQIQTGKPLPAGTILKLVTSADGKPTTI 735

QY 215 -----EMGNVKKVLBOLLDAKIEAEFFRKLYVELKSSPQHLVPLPKSVVALRQLLP 269
Db 736 ITTQASGAGTKPTILGI-----SSVSPGTT---KPGTTTIKTIP 773
QY 270 NSOSFTQCCVQQTSSDMVIATCTTTVT-----SPVVT-----TTVSSSQSEKSI 314
Db 774 MSALITQAGATGVTSPGKISPIITIIKVTSGTGAPAKIITAVPKIATGHGQQGVTV 833
QY 315 IVSGA-----TAPRT-----VSVQTLNPLAGPVGAK--AGVVTLHRSVGPAAATG 356
Db 834 VLKAGAPQPCITILRTVPMGGVRLVTPVTSVAVKPAVTTLVKGTGTTGLTGTGTST- 892
QY 357 GTTAGTGLLTPLVTSVA--NVTVTSLQPEKPVVSGTAVTSLSPAVTFTGETSGAIC 414
Db 893 -SLAGAGGHSASLASLPTITGLTIATLSQ-----VINPTAITVSAQTTLTAAGGLTTP 947
QY 415 LPVVKPVVSCWDHICKPVGTGVQIKL-AQPGVLSQPA-GIPTGSSSKQLSLFHVVQ 472
Db 948 TITWQPV-----SQPTQVTLITAPSGVEAQPVDLPVS-----ILA 983
QY 473 QPSGGNEKQVTTISHSTLTIIQKCGQKTMFVNTIIPTSQPPASI-----LKQITLPG 525
Db 984 SP-----TTEQPTATVTIADSGQGVQPGVTLVCSNPPCETHETGTNTATTIVA 1035
QY 526 NKILSLQASPTQRNRIKENVTSFRDEDDINDVTSMAGVNLEENACILATNSLVGTLI 585
Db 1036 N--LGGHPQPTQVQFV-----CDRQEAASLVSTVG-----QON-----GSV 1072
QY 586 QSCKDEP 592
Db 1073 RVCNPP 1079
RESULT 5
HCFCL_MESAU STANDARD; PRT; 2090 AA.
ID HCFCL_MESAU
AC F51611;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Host cell factor C1 (HCF) (VP16 accessory protein) (HFC1) (VCAF) (CFF)
GN HCFCL.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RA Goto H., Nishitani H., Umene K.I., Nakabepu Y., Nishimoto T.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: UPON LYTIC INFECTION OF PERMISSIVE CELLS, THE HSV
CC TRANSACTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF
CC ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING
CC PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR
CC ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES (BY
CC SIMILARITY).
CC -1- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM
CC 110 TO 150 kDa AND A MINOR 300 kDa POLYPEPTIDE. THE MAJORITY OF N-
CC AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBEIT
CC NONCOVALENTLY, ASSOCIATED (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE
CC SIGNAL (BY SIMILARITY).
CC -1- PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH
CC CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS
CC AT A DEFINED SITE, PPCE/THET, WITHIN THE HCF REPEAT (BY
CC SIMILARITY).
CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; D45419; BAA08258.1; -
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR001798; Kelch.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF01344; Kelch; 5.
DR SMART; SM00060; FN3; 2.
KW Nuclear protein; Repeat.
FT REPEAT 44 89
FT REPEAT 93 140
FT REPEAT 148 194
FT REPEAT 217 265
FT REPEAT 266 313
FT DOMAIN 1010 1448
FT REPEAT 1010 1035
FT REPEAT 1010 1035
FT REPEAT 1072 1097
FT REPEAT 1101 1126
FT REPEAT 1157 1182
FT REPEAT 1295 1320
FT REPEAT 1323 1348
FT REPEAT 1358 1383
FT REPEAT 1423 1448
SQ SEQUENCE 2090 AA; E495E8B1F2385E17 CRC64;

Query Match 5.6%; Score 238; DB 1; Length 2090;
Best Local Similarity 21.7%; Pred. No. 0.00012;
Matches 147; Conservative 84; Mismatches 219; Indels 226; Gaps 32;

QY 10 VSAPKRVSSGRLPAQIVAVKAPNTTIIQFPANLQPLPGTVLIKNSGPMVLSPQQT 69
DB 537 IGSNFMQSMGMAAALAAATQKIPSSA-----PTVLVSFAGTIVT-----VAVTPGTT 588
QY 70 TRAEITSNITSRPVNPQPT-----VKICTVPNSSQ-----LIKKVAVTPVKKL 115
DB 589 LPA--TVKVASSPMVSNPATRMLATAAAQVGTSSSAANTSTRILIVHKSGETVVAQ 646
QY 116 AQIGTVV-----TVPKP-----SSVOSVAVPTS----- 140
DB 647 AQVTVTVGGVTKTITLVKSPISVPGSALISNLGKMSVQTKPVQTSVATGQASTGPV 706
QY 141 -----VVTVTPGRPLNTVTL-----KPSLSGAST-----PSNE 170
DB 707 TQIIQTKGPLPAGTILKLVTSADGRTTIIITQASGAGSRPTILIGISVSPTSSTTKPGTT 766
QY 171 PNLKAENSAVQINLSPTMLNKKCKNFKLAML-IKLACSGSQSPGEMGQNVKKLVEQLLD 229
DB 767 TLIITPMSAIIQAGATGVTSTPGIKSPITITTKVMTSGTGAP-----AKIITAV-- 818
QY 230 AKIEAEETRKL-YVELKSSP-QPHLVFELKKSVALHQLPLNSOSFIOQCQVQQTSSDMV 287
DB 819 PKIATGHGQGGQTVVLKAGPQPG-----AILRTVP-----MSGVR 855
QY 288 IATCTTTTTSVPTVVTTSSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKGVTLH 347
DB 856 LVTPVTVSAKPAVTT-----LVVKGTT-----GVTTIG 884
QY 348 SVGPTAATGGTAGTGLLOTSKPLVTSVA--NTVTVLSLQPEKPVVSGTAVTSLSPAVTF 405
DB 885 TVTGTVST--SLAGAGHASTASLATPITTLGTIATLSQ-----VINTAIVTAAQTTL 938
QY 406 GETSAAICLPSVKPVVSCDWHICKPVIGPVQKL-AQGPVLVSQPA-GIPTGSSSKQ 463
DB 939 TAAGGLTTPTITMQPV-----SQPTQVTLITAPSGVEAQPVHDLPSV----- 980
QY 464 LPSFLHVVOQSGGNEKQVTTISHSSTLTIOCKGOKTVPVNTIIPTSQFPFASI----- 517
DB 981 -----ILASP-----TTEQPTATVTIADSGGQVQPTVTLVCSNPPCETHETGTT 1026

QY 518 -LKQITLPGNKILSLQASPTQKNRIKENVTSCFRDEDDINDVTSAGVNLNENACILAT 576
DB 1027 NTATTIVVAN--LGGHPQPTQVQFV-----CDRQEAASLVTSVAG-----QQN----- 1068
QY 577 NSELVGTLIQSKDEP 592
DB 1069 -----GNVVRVCSNPP 1079

RESULT 6

AMYL_YEAST
ID AMYL_YEAST STANDARD; PRT; 1367 AA.
AC P08640; P08068;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucoamylase SI/52 precursor (EC 3.2.1.3) (Glucon 1,4-alpha-
DE glucosidase) (1,4-alpha-D-glucan glucohydrolase).
GN STAL OR STA2 OR MAL5 OR YI019C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Barrell B.G., Badcock K., Bankier A.T., Bowden S., Brown D.,
RA Chenters C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagers K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX MEDLINE=87194600; PubMed=3106330;
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
RT STAL".
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STA2 and SGA genes
RT from Saccharomyces cerevisiae".
RL FEBS Lett. 239:179-184(1988).
CC -|- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -|- SIMILARITY: TO S.POMBE SPBC215.13.
CC -|- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z38061; CAA86176.1; -
DR EMBL; M16164; AAA35014.1; -
DR EMBL; M16165; AAA35015.1; -
DR EMBL; X13857; CAA32069.1; -
DR PIR; B26877; B26877.
DR PIR; A26877; A26877.
DR PIR; S48478; S48478.
DR SGD; S0001458; MUC1.
KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.

Db 803 PTEETISTEVTVSPDETTLPTEVTVSTEVTVNS-----PETSVPPEETI----- 849
QY 339 AKAGVVTLHSVGTAAAGTTAGTGLL-----QTSKPLVTSVANTVTVTS-----LQPEK 388
Db 850 -----LTLTVTEVTPVTEVTVGHTVTVNSPEETSVPTEETISTEVTVSPDETTLPTEV 905
QY 389 PVSSTAVTSLPAPVTFGETSGAAICLSVKP-----VVSFWDHICKVIGTPVQIK 441
Db 906 PTVS-TEVTVNSPEET-----SVPEETILTETTVSPDETVFPTEGTTL----- 949
QY 442 LAQPGVLSQAGI---PTGSSSKOLFSLFHVQVQPSGNEKQVTTISHSS----- 490
Db 950 ---PTEVLTVPIEVTFPGETT---VPTVTVSTVSTVTEVTVSTVTVPEETSITVETAT 1004
QY 491 -----LTIQKCGKQTMVNV-TIPTSQFPP---ASILKQIPLGNKILS 530
Db 1005 VLPASIPPEETTPTEVTTTPPEETTPAEVTVTPPASIPPEETASLTVETITTPPEETT 1064
QY 531 ---LQASPTQKNRIKENVT-----SCFDEDDI-NDVTSMAQ---VNLNEENACILATNSE 579
Db 1065 PTEVTVTPPEKTITPTEVTVTPPASIPPEETTVPEETTPPEETTPPEETTVSTQETTLTQESA 1124
QY 580 LVGTLLI---QSCKDBPFLFAGL 599
Db 1125 VTQTSIACRPPCPSPPLMPIGPL 1147
RESULT 9
CPN_DROME STANDARD; PRT; 865 AA.
AC Q02910;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-FEB-1994 (Rel. 27, Last sequence update)
DE Calphotin.
GN CPN OR CAP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=93165729; PubMed=8094559;
RA Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
RT "Calphotin: a Drosophila photoreceptor cell calcium-binding protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=93165730; PubMed=8434015;
RA Ballinger D.G., Xue N., Harshman K.D.;
RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds
RT calcium and contains a leucine zipper.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
CC -1- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO
CC REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL
CC OF CA+2 PER MOL OF PROTEIN.
CC -1- SUBUNIT: HOMODIMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
CC -1- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
CC COMPOUND EYES AND OCELLI.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
CC DEVELOPMENT.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----L02111; AAA28405.1; -
DR EMBL; L02111; AAA28405.1; -
DR PIR; A47282; A47282
DR FlyBase; FBgn0010218; Cpn.
KW Calcium-binding.
FT CONFLICT 36 36 A -> AVAPAVVA (IN REF. 2).
FT CONFLICT 43 43 I -> T (IN REF. 2).
FT CONFLICT 64 64 I -> V (IN REF. 2).
FT CONFLICT 76 76 T -> A (IN REF. 2).
FT CONFLICT 100 100 P -> PP (IN REF. 2).
FT CONFLICT 126 127 VO -> AP (IN REF. 2).
FT CONFLICT 154 154 I -> V (IN REF. 2).
FT CONFLICT 160 160 S -> T (IN REF. 2).
FT CONFLICT 534 534 A -> E (IN REF. 2).
FT CONFLICT 699 699 I -> T (IN REF. 2).
FT CONFLICT 703 703 V -> L (IN REF. 2).
FT CONFLICT 721 721 D -> E (IN REF. 2).
SQ SEQUENCE 865 AA; 84781 MW; 2110417E0B0E7CFE CRC64;
Query Match 4.8%; Score 206.5; DB 1; Length 865;
Best Local Similarity 21.7%; Pred No. 0.0015;
Matches 122; Conservative 82; Mismatches 242; Indels 115; Gaps 22;
QY 1 GTLVTKV-APVSAP-----PKVSSGRLRPAP-----QIVAVKAPNTTTIQFPANLQ 45
Db 4 GTIPSPVAPVAAVPTPSAAPAAPVQVVAAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 63
QY 46 LPPTVILKSNGLMLYSPOQTVTRAETTSNITSRPAVPANPQVTKICTVPNS-SSQLI 104
Db 64 IPAPAPIAAASVTPVASVAP---PVVAAATP-----PAASPVSTPVAQAIDVAVSAPVA 115
QY 105 KKVAVTPVKKL-AOIGTTVTVTVPKSSVQSVAVPTSVVT-VTPGKPLNTVTLKPSL 162
Db 116 PVAATPTPVQIIPVAAPVIAATPPVAASAPTAAVTPVISPVIAFPVVPVANTTPVVAAP 175
QY 163 ASSTPSNEPNLKAENSAVQINLSPTMLE-----NVKCKNFLAMLIKACS 209
Db 176 VAAVPAAPVVAAPVLAAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAP 227
QY 210 GSQSPMGONVKKLYEQLLDADKAEAEETKRLYVELKSSQPHLPFLKKSVALRQLLP 269
Db 228 -----PEVSVVATKPLAAAEPPVVAAPPATETPVVAPAAASPHVSVAPAVETAVVA 278
QY 270 NSQSFIOQCQVQOTSSDMVIACTTTTTPSVVTTTSSQSEKSIIVSGATAPRTVSVOT 329
Db 279 VSAS-----TEPPVAAATLTAPET-PALAPVVAESQVAANTVATPPTP-APEPET 328
QY 330 LNPAGPVGAKAGVTVLHSGVGTAAATGGTTAGTLQTSKPLVTSVANTVTVTSLOPEKP 389
Db 329 IAP-----PVAETPEVASVAVAEETPPVVPVVA-----AESIPAP 364
QY 390 VVSGTAVTSLPAPVTFGETSGAAICLPSVKPVVS-----FCWDHICKPVIGT 436
Db 365 VVATTPVPATL-AVTDPDVDTASV--PELPPVIAAPSPVPSAVAETPVDLAPVLPVAAE 421
QY 437 PVQIKLAPGVPVLSQAGIPTGSSSKOLFSLFHVQVQPSGNEKQVTTISHSSSTLTIOK 496
Db 422 PVPVAVAEETETPAPASAPVITIALDPEVAPVIAAASDAPAE-----APSAAPAI 473
QY 497 GQKTMVNTIPTSQFPPASI 517
Db 474 -VSTPTTASVPETTAPPAV 493
RESULT 10
DAN4_YEAST STANDARD; PRT; 1161 AA.
AC P47179;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cell wall protein DAN4 precursor.

FT	CARBOHYD	464	464	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	509	509	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	554	554	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	599	599	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	644	644	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	689	689	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	734	734	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	888	888	N-LINKED (GLCNAC. . .)	(POTENTIAL).
SQ	SEQUENCE	1322 AA;	138072 MW;	AAPDF1FF13267CEA	CRC64;
Query Match					
Best Local Similarity 4.8%; Score 203.5; DB 1; Length 1322;					
Matches 142; Conservative 114; Mismatches 276; Indels 215; Gaps 31;					
QY	22	LPAPQ-IVAKAPNTTITQF	-----PANLQLPFGTGLIKNSG	-----PLMLVSPQOT	68
DB	663	LPTDETIIVIRPTATTAMTTQPMNDTFTSTSTETITVGTNGLPTDETIIVIRPTT	722		
QY	69	VTRAETT	-----SNITSRAPVNPQNTKICTVPNSSOLIKKVAIVPVK	113	
DB	723	ATTAMTTQPMNDTFTSTSTETITVGTNGLPTDETIIVIRPTTSE	768		
QY	114	KLAQIGTTVTVTPKPSVQSVAVPTVVVTPGKPLN-TVTLK-PSLGLASSTPSNEP	171		
DB	769	-----GLISTTPTWGTFTSTSTETITVGTNGLPTDETIIVIRPTTSE	822		
QY	172	-----NLKAENSAVQINLSPMLNENVKCKNFAMLIK	222		
DB	823	WTGTFSTSTETITVGTNGLPTDETIIVIRPTTSE	869		
QY	223	LVEQLLDKAKIEBEETRKLYELKSSPOHVLFPFKKSVVALRQLLPNSQFIOCVQOT	282		
DB	870	-----ITSFITSARPIITPFYPS	900		
QY	283	S--SDMVIATCTT--VTTSPVTTTSSQSEKSIIVSGATAPRTVSQTLNPLAGPVG	338		
DB	901	SDTSLVSSSVTSLVSSPVISSFFISP	946		
QY	339	AKAGVTVTLHSGVPTAATGTTAGTGLLQTSKPLVTSVA--NTVTVSLQPEKPVVSGTAV	396		
DB	947	SKSSVIPTSS--STGSGSESETGSASSASSSSSSISSESPKSTYSSSLPPVTSATTSQEI	1004		
QY	397	TLSLPAVTFEGETSGAICLPVKPVVFCWDHICKPVGTGPV---QIKLAQPGVLSQP	452		
DB	1005	TSSLPPVTTTSEQTTLV---TVTSCSHVCTESISSAIVSTATVTVSGATTEYTTW	1059		
QY	453	AGIPTGSSSKOLFSLFHVVOOPSGNGEKQVTTISHSLLTIQKC	498		
DB	1060	CPISTETITKQ---TTETTKQTKGTETTTETTKQTTVVTTISSCESDVCSKTASPAIVST	1116		
QY	499	KTMPVNTI	549		
DB	1117	STATINGVTTETVTCPISTTESKQOTLVTVTSCGVCSETTSPA	1170		
QY	550	RDEDDINDVTSWAGVNLNEENACILATNSLVGLIQCKDEPFLFIALQ	600		
DB	1171	-----TVNDVTVTVISWRPQ	1220		
QY	601	-----KRLLDGGKHDITELNSDAVNLSQATQRL	631		
DB	1221	TKTVTVTSSISRENHAETASATDVIGHSSSVSVSETGNTKSLTSSGLSTMSQ	1278		
QY	632	RGLEKL-TATAQHRMTTYKASENYIL	657		
DB	1279	TPASSMVGSSASLEISTYAGSANSLL	1305		
RESULT 12					
AC	AGAL_YEAST	STANDARD;	PRT;	725 AA.	
AC	P32323;				
DT	01-OCT-1993 (Rel. 27, Created)				
DT	01-OCT-1993 (Rel. 27, Last sequence update)				

DT	01-OCT-1996 (Rel. 34, Last annotation update)				
DE	A-agglutinin attachment subunit precursor.				
GN	AGAL OR YNR044W OR N3431.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
OX	NCBI_TaxID=4932;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=91304412; PubMed=2072914;				
RA	Roy A., Lu C.F., Marykwas D.L., Lipke P.N., Kurjan J.;				
RT	"The AGAL product is involved in cell surface attachment of the				
RT	Saccharomyces cerevisiae cell adhesion glycoprotein a-agglutinin.";				
RL	Mol. Cell. Biol. 11:4196-4206(1991).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RL	Pohl T.M.;				
RA	Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.				
CC	-!- FUNCTION: MEDIATES THE CELL SURFACE ATTACHMENT OF THE A-AGGLUTININ				
CC	SUBUNIT. S.CEREVISIAE A AND ALPHA CELLS EXPRESS THE COMPLEMENTARY				
CC	CELL SURFACE GLYCOPROTEINS A-AGGLUTININ AND ALPHA-AGGLUTININ,				
CC	RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO PROMOTE CELLULAR				
CC	AGGREGATION DURING MATING.				
CC	-!- SUBUNIT: CONTAINS AT LEAST A BINDING SUBUNIT DISULFIDE-LINKED TO				
CC	A CORE SUBUNIT.				
CC	-!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR				
CC	AND/OR BY CARBOHYDRATE-MEDIATED COVALENT CROSS-LINKS (POSSIBLE).				
CC	-!- PTM: EXTENSIVELY O-GLYCOSYLATED.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; M60590; AAA34382.1; -				
DR	EMBL; 271659; CAA96325.1; -				
DR	PIR; S17031; S17031.				
DR	PIR; A41258; A41258.				
DR	SGD; S0003327; AGAL.				
KW	Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat;				
FT	Pheromone response.				
FT	SIGNAL	1	22	POTENTIAL.	
FT	CHAIN	23	725	A-AGGLUTININ ATTACHMENT SUBUNIT.	
FT	DOMAIN	53	493	2 X APPROXIMATE REPEATS.	
FT	REPEAT	53	149	1-1.	
FT	REPEAT	395	493	1-2.	
FT	DOMAIN	182	307	18 X APPROXIMATE TANDEM REPEATS, SER/THR-	
FT	REPEAT	182	188	RICH.	
FT	REPEAT	189	195	2-1.	
FT	REPEAT	196	202	2-2.	
FT	REPEAT	203	209	2-3.	
FT	REPEAT	210	216	2-4.	
FT	REPEAT	217	223	2-5.	
FT	REPEAT	224	230	2-6.	
FT	REPEAT	231	237	2-7.	
FT	REPEAT	238	244	2-8.	
FT	REPEAT	245	251	2-9.	
FT	REPEAT	252	258	2-10.	
FT	REPEAT	259	265	2-11.	
FT	REPEAT	266	272	2-12.	
FT	REPEAT	273	279	2-13.	
FT	REPEAT	280	286	2-14.	
FT	REPEAT	287	293	2-15.	
FT	REPEAT	294	300	2-16.	
FT	REPEAT	301	307	2-17.	
SQ	SEQUENCE	725 AA;	73353 MW;	70420C853B0B01F8	CRC64;

Query Match 4.7%; Score 201; DB 1; Length 725;
Best Local Similarity 20.4%; Pred. No. 0.0023;

Matches 144; Conservative 120; Mismatches 309; Indels 134; Gaps 30;

```
QY 4 VTKVAPVAPKVSQGRPLPAQIVAVKAPN-----TTTIQPPANIQ-----LPPGTGLI 53
DB 18 LTNIALASDPETI-----LVTITKNDANGVVVTTTTS-PALVSTSTIVQAGTTTL 66
QY 54 KNSGGPLML-----VSPO-----QTVRAET--TSNITSRPAVPANP--QTVKICTVP 97
DB 67 YTWCPPLTVTSAAELSPISVATLSRFSTLTLSVECHACFSSSLPPTTLTUSVTS 126
QY 98 NSSQLIKKVVAVPVKLAQIGTVV--TTVPKPSVQSVAVPTSVV--TVTPGKPLMTVT 154
DB 127 KFTSYICPTCHTTAISLSLSEVGTTVVSSSAIEPSSAIIISPVTLSLSTSNP--TTT 184
QY 155 TLPPSSLCASSTPSNPNLKAENSAVOINLSPMLNFKCNFLAMLKLCAGSSQSP 214
DB 185 SLSTSTSPSST-STGPSSTSTSSSTSTSSSTSTSTSTSTSTSTSTSTSTSTSTST 228
QY 215 EMGQNVKLVQLLDKIAEETRLKLYVELKSSPQPHLVFLKKSVVALLRQLLPNSOF 274
DB 229 SSTSTSSSLTSTSSSTSTSTSSSTSTSTSTSTSTSTSTSTSTSTSTSTSPSKST 282
QY 275 IOOCVOOTS-----SDMVIATCTTTVTTPVTTTSSSQSEKSIIVSGATAPRTVSQVT 329
DB 283 SASSTSTSYSTSTSPSLTSSPTLASTSP-SSTSISSSTFTDSTSSLGSSIASSTSVSL 341
QY 330 LNPAGPVGAKAGVVLHSGVPAATGCTTAGCGLLOTSPKPLVT--SVANTVTTVLSQPE 387
DB 342 YSP-STPVSVPS--TSSNVATPSMTSSVETTVSSQSSEYITPKSISITIPFSMSY 398
QY 388 KPVVSG--TAVTSLPAVTEGSGAICLPVSKPVVFCWDHICKPVICTPVQIKLAQP 445
DB 399 FTVVSGVTTWYTCVPYSSESESTSTLSMHEVTTTATVCTHESCMP-----445
QY 446 GPVLSPAGIPTGSSSKQLFLFHVQOPSGGNEKQVTTTISHSTLTIOKCGKTPMPT 505
DB 446 ----SQTSLTSS-----IKMSTKNVATSVSTSVSSVACSTCAETSHSYSS 490
QY 506 IIPTSOPFPASILKQTLPGNKILSQASPTQKNRI---KENTVTSFRD--EDDINDVTS 560
DB 491 VQPASS---SSVTQOTSTKSWSSMTSDENFKHATGKYHYTSSOTSTISVSSEATS 547
QY 561 MAGVN--LNEENACILAT-----NSELVGTLIQCKDEPFLFIALOKRILDKKKHIDTE 614
DB 548 TSSIDESQSQSHSLSTSVLSSSLSATL---SSDSTILLFSSVSSLSVE---QSPVTT 601
QY 615 LNSDAVNLISQAQTERLGLLEKLTATAQHRMTTYKASENYILCSDT 661
DB 602 LQISSTSEILQPTSS-----TATATISASTSLSATSISTPST 639
```

RESULT 13

```
ZAN_HUMAN STANDARD; PRT; 2700 AA.
ID ZAN_HUMAN
AC Q9Y493; O00218;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zonadhesin (Fragment).
GN ZAN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=99018118; PubMed=9799793;
RA Glockner G., Scherer S., Schattevoy R., Boright A., Weber J.,
RA Tsui L.C., Rosenthal A.;
RT "Large-scale sequencing of two regions in human chromosome 7q22:
RT analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci
RT reveals 17 genes."
RL Genome Res. 8:1060-1073(1998).
```

```
[2]
RN SEQUENCE OF 2338-2700 FROM N.A.
RP TISSUE=Testis;
RC MEDLINE=97271566; PubMed=9126492;
RA Gao Z., Harumi T., Garbers D.L.;
RT "Chromosome localization of the mouse zonadhesin gene and the human
RT zonadhesin gene (ZAN).";
RL Genomics 41:119-122(1997).
CC -!- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC SIGNALING.
CC -!- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.
CC -!- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
CC ZONA PELLUCIDA.
CC -!- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHUS.
CC -!- DOMAIN: THE VWFD DOMAINS 2 AND 3 MAY MEDIATE COVALENT
CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
CC -!- SIMILARITY: CONTAINS 3.5 MAM DOMAINS.
CC -!- SIMILARITY: CONTAINS 4.5 VWFD DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
```

```
FT CARBOHYD 2505 2505 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CONFLICT 2374 2379 NNQKMA -> RAGPGP (IN REF. 1).
FT NON_TER 2700 2700
SQ SEQUENCE 2700 AA; 293013 MW; 80E60CC0B12277B1 CRC64;

Query Match
Best Local Similarity 4.7%; Score 199; DB 1; Length 2700;
Matches 147; Conservative 91; Mismatches 268; Indels 160; Gaps 34;

QY 1 GTLVTKV--APVSAPP-----KVSSGPRLPAPQIV-AVKAPNTTII 38
DB 372 GTMLELLGSPAGSPGPIPLKRVGSRPYQWNTSVTPSGHQWQLFKGQGSTASV 431
QY 39 QPANQLPDPGVLKNSGPMVLVSPQOTVTRAEITSNITSRPVPANPQVACTVFN 98
DB 432 VAMGFILINPGCPKVKV-LPELPPVSPVSGTGPSETT-GLTENPTISTKKTPTS 483
QY 99 SSSQLIKKVAIVPVKLAQIGTVTTVPKPSVQSVAVPTSVVIVTPCKPLNTVTLKP 158
DB 484 -----IEKPSVTEK-----PTVPK-----EKPTIPEKPTISTEKP--TIPSEKP 522
QY 159 SSGASSTPSNPNLKAENSAVAQINLSPTMLENVKCKKNFLAMLIKACSGSQSPMGQ 218
DB 523 -----NMPSEKPTIPSE-----KPTILTPEKTPIS-----EKPTIPSEKPTI-- 559
QY 219 NVKLVQELLDKABEEETRKLYVELKSSPOPHLVFLKKSVALROLPLNSQSFY--Q 276
DB 560 STEKPTVTEPTTPTTEET--TYME-----EP-VIPEKPSIPEKPSIPEKPTISME 611
QY 277 QCVQOTSSDMVIATCTTTVTTSPPVTTTVSSSQSEKSIIVSG-----ATAPRTVS-----V 327
DB 612 EYIISTEKPTICEKPTIPEKPTIPEKPTISTEKSTISPEKPTTPEKPTIPEKPTISTEKP 671
QY 328 QTLNLAGP--YKAGAGVVTLSVGTGAATGTTAGTGLLOTSKPLVTSVANTVTT--VS 383
DB 672 PTEKPTISPEKLTIPTEKLTIPTEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTI 726
QY 384 LOPEKPVSGTAVTLSLPAVTGETSGAAICLPSPVKPVVSCFWDHICKPVIGT-----P 437
DB 727 IPMEKPTLPTTEETTTTSVEETHT--STEKLTIPMEKPTIS-----TEKPTIPEKPTISP 778
QY 438 VOI-----KLAQPG-----PV-----LSQAPAGITGSSSKQLSFLFHVQVQPSGN 478
DB 779 EKLTIPTEKLTIPTEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTI 834
QY 479 EK-----QVTTSHSSTLTIOCKGQKTMVNTIIP-----SQFPPASILKQITLPGNKI 528
DB 835 EKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEK 892
QY 529 LSQASPTQKNRIKENVTSCFDEDDINDVTSMAGVNLEE-----NACILATNSLVG 582
DB 893 ----PTIPEKPTIPEKLTALRPPHPSPPTAGLAAALVMSHPAPSPMTSVILGTTTTSRS 949
QY 583 TLIOQC 588
DB 950 STGMSC 955

RESULT 14
YJH8_YEAST
ID YJH8_YEAST STANDARD; PRT; 881 AA.
AC P47033;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 89.2 kDa protein in SCPl60-SMC3 intergenic region.
GN YJL078C OR J1027
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN-S288c;
RX MEDLINE-96093911; PubMed-7483841;
RA Miosga T., Schaff-Gerstenschlaeger I., Chalwatzis N., Baur A.,
RA Boles E., Fournier C., Schmitt S., Velten C., Wilhelm N.,
RA Zimmermann F.K.;
RT "Sequence analysis of a 33.1 kb fragment from the left arm of
RT Saccharomyces cerevisiae chromosome X, including putative proteins
RT with leucine zippers, a fungal Zn(II)-Cys6 binuclear cluster domain
RT and a putative alpha 2-SCB-alpha 2 binding site.";
RL Yeast 11:681-689(1995).
CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: Z49353; CAA89370.1; -
DR EMBL: X83502; CAA58492.1; -
DR EMBL: X88851; CAA61314.1; -
DR HSP; P04284; ICPE.
DR SGD; S0003614; YJL078C.
DR InterPro; IPR001283; Allrgn_V5/Tpx1.
DR PRINTS; PR00188; SCP; 1.
DR ProDom; P000542; Allrgn_V5/Tpx1; 1.
DR SMART; SM00198; SCP; 1.
DR PROSITE; PS01009; SCP_AG5_PRL_SC7_1; 1.
DR PROSITE; PS01010; SCP_AG5_PRL_SC7_2; 1.
KW Hypothetical protein.
FT DOMAIN 166 379 ALA/SER/THR-RICH.
SQ SEQUENCE 881 AA; 89152 MW; 37DBAC660CA9D12A CRC64;

Query Match
Best Local Similarity 4.6%; Score 197.5; DB 1; Length 881;
Matches 145; Conservative 87; Mismatches 268; Indels 189; Gaps 27;

QY 47 PGTVL-----IKNSGPMVLVSPQOTVTRAEITSNITS--RPVAPNPQV 91
DB 146 PPGNLTGFAEEVEPLISTVSSSS-----SSSTSTSDTSTVSTISSIMPV-AQGYTT 199
QY 92 KICTVPNSSOLIKKVAIVPVK--KLAQIGTVVTVTPKP-----SSVQSVAVPTSVVT 143
DB 200 TVSSAASSSS--LKSTTINPAKTAITLASSSVITTSSTESVGSSTVSSASSSVTTSYAT 257
QY 144 ---VTPGKPLNTVTLKPSLGSASSTPSNEPNLK-----AENSAVQINLSPTMLE 191
DB 258 SSSTVVSDDATSTTTTSSVATSSSTSSDPTSSATAASSSDPSSAAAS---SSASTE 314
QY 192 NVYKKNFLA-----MLIKLACSGSQSPMGQNVK-----KLVEQ 226
DB 315 NAASSSSAISSSSSMVSAPLSLTLTSTASSRSVTSNSVNVKFAANTTVFSAQTTSSVA 374
QY 227 LLDKIAEEETRKLYVELKSSPQPH--LVFPLKKSVVALROLLPNSQS-----FIOQCQV 280
DB 375 SLSSVAADDIOGSTSKATSVSEHTSIVTSATNAQYATRLGSSRSSGAVSSSVA 434
QY 281 QTSSDMVIA-----TCITTVTTSPPVT--TTVSSSQSEKSIIVSQA 319
DB 435 QSVLNSVIAVNTDVSVTSVSSSTAHTTKDTATTSVTAASEITSETAQASSTSEKINSAA 494
QY 320 TAPRTVSVQTLNPLAGPVGAKAGVTLHSVGTATGTTAGTGLLOTSKP-----370
DB 495 T-----SSSIYSNASVSGHVTYAAEYAITSEQSALATSPATNCSSIVK 541
QY 371 -----LVTSVANTVTVSLQPEKPVVSGTAVTSLPAV-----403
DB 542 TTLENSSTTITATTKSTTILATTANNSTRAATAVTD-PTLDPDTSASPTDNKHTS 600
```

```

QY 404 TFGETSGAAICLPVSKPVVSPFCMDHICKPVIGTPVQIKLAQGPVL----- 449
Db 601 TYGSSSTGCA-SLDSLRRTTTSISVSNTQVSTCTSDSDSPFALSTATTESNLIT 659
QY 450 -----SQAPAGITPGSSKOLFSLFHVVOQPSGGNEKQVTTISHSSTLTITQKCGQKTMP 502
Db 660 NTITASCSTDSNFTSAASSTDETAFTRTISTS-----CSTLNGASTQT-SELTITSPMK 712
QY 503 VNTIIPSTQFPPASILQKITPLGKNKILSLQASPTQKNRI-----KENVTSQCFRDEDDI 555
Db 713 TMTVVPASSFPSTT-----TTCLENDNDTAFSSIXTEVNAATLIINPGETSSLASDASFATSEKP 768
QY 556 NDVTSMAGVNLNEENACILATNSLSELVGL 584
Db 769 NEPTSVKSTS-NEGTSSTTTTYQQTVA TL 796

RESULT 15
FIG2 YEAST
ID FIG2.YEAST STANDARD; PRT; 1609 AA.
AC P25653;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Factor induced gene 2.
GN FIG2 OR YCR089W OR YCR89W OR YCR1102.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92397594; PubMed=1523889;
RA Wilson C.; Grisanti P.; Frontali L.;
RT "The complete sequence of a 6146 bp fragment of Saccharomyces
RL cerevisiae chromosome III contains two new open reading frames.";
RL Yeast 8:569-575(1992).
CC -1- FUNCTION: REQUIRED FOR EFFICIENT MATING.
CC -1- INDUCTION: BY MATING PHEROMONES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59720; CAA42254.1; -.
DR PIR; S19504; S19504.
DR PIR; S25345; S25345.
DR SGD; S0000685; FIG2.
SQ SEQUENCE 1609 AA; 166049 MW; 7066AD7F85A7B852 CRC64;

Query Match 4.6%; Score 197.5; DB 1; Length 1609;
Best Local Similarity 21.1%; Pred. No. 0.01;
Matches 166; Conservative 106; Mismatches 327; Indels 187; Gaps 34;

QY 2 TLVTKVAPVAP-PAKVSQPRPAPQIVAVKAPNTTTIQPPANLQPLPPGVLLKNSGPL 60
Db 799 TVITSVTATSTIPLLSSTSVLSSTVSEGAKNPAASEVTINQV-SATSEATSTSTQV 857
QY 61 MLVSPQOVTVAETTSNITSRPVPANPQVTKITV-----PNSSQLIKKVAVTPVKKL 115
Db 858 SATSATATASESSTTSQVSTASETISTLGTQNTFTTGLLFPALSTEMINTVVS--RKT 915
QY 116 AQIGTTVVTVTPKPSVQSVAVPSTVVTVTPGKPLNTVTLKPSLSGASSTPSNEPLKA 175
Db 916 LIISTEVCS-----HSCKVPVITEVTSK-----GPPSNHSSQT 951
QY 176 ENSAAVQINLS--PTMLENVKCKKNFL---AMLIKACSGSQSPM-----GQNVKKLVE 225
Db 952 LQTEAVEVTLSSHQVTMTSEVCSNICTPTVITVSQMRSTPPFPYLSTSSSSSLASTKK 1011

```

```

QY 226 QLLDAKIEAEFEKRLKVELKSSPOPHLPPL-----KKSVALROLPLNPSQSFIOQCVOQT 282
Db 1012 SSLEASSEMSTFS-----VSTQSLP-----LAFTCEKRSTTSVSQW---SNTVLNTIMSS 1060
QY 283 SSDMVIATCTTIVTTS-----VTTTVSSSQ-----SEK 312
Db 1061 SSVINISNEKPSSTSPYNFSSGYSLPSSSTPSPSYSLSTATTINGIKTVTTWCPLAEK 1120
QY 313 SIIVSGATAPRTV-----SVO-TLNPLAGPVGAKAGVVTUHLHVGPTAAT 355
Db 1121 STVAASSQSSRSVDRFVSSSKPSSLSQTSIQYTLSTATTISGLKTVTTWCPLTSKST 1180
QY 356 GGTAGTAGLLOTSKPLVTSVAN-TVTVTSLOP-----EKPVVSGTAVTLSPAV 403
Db 1181 LGATTQTS--STAKVITSASSATSTISLSTSESSSSGYLSKGVCSGTECTQDVPTQ 1238
QY 404 TFGETSGAAICLPVSKPVVSPFCMD-----HICKPVIGTPVQIKLAQGP--PVLSQ 451
Db 1239 SSSPASTILAYS-PSVSTSSSSSFSTTTASTLTSTHTSVPLLPSSSSISASSPSTSLST 1297
QY 452 PAGIPTSSSKOLFSLFHVVOQPSGGNEKQVTTISHSSTLTITQKCGQKTMPTVITIPTSQ 511
Db 1298 SLSPAFSTSS-----TLPTATAVSSSTFIASSLPLS-SKSSLSLSPVSSSILMSQ 1346
QY 512 FPPASILKQ--ITLPGNKILSLQASPTQKN-RIKENVTS-----CFRDEDDI---- 555
Db 1347 FSSSSSSSSSLASLP-----SLSISPTVDTVSVLQPTTSIATLTCTDSQCOQEVSTICNG 1401
QY 556 ---NDVTSMAGVNLNEENACILATNSLSELVGLIOSCKDEPFLFGLAKLRILDIG----- 607
Db 1402 SNCDVTSATTPPSTVTDVMTCTGSECQKTSSEC-DGYSCKVSEYTKSSATISACSGE 1460
QY 608 --KKHDITELNSDAVNLISQATQERLRL-----LEKLTAAQHRMTTYKASENYILCS 659
Db 1461 GCOASATSELNSQVVTMTSVITPSAITTSVEVHSTESTISITTVKPVYTS-----S 1513
QY 660 DTRSQL 665
Db 1514 DTNGEL 1519

```

Search completed: February 16, 2003, 21:55:46
Job time : 42.4197 secs

This Page Blank (uspto)

GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:59:29 ; Search time 21.8929 Seconds
(without alignments)
1145.042 Million cell updates/sec

Title: US-09-763-909-2

Perfect score: 4264

Sequence: 1 GTLVTKVAPVSAPPKVSSGP.....KZHQHERALFTIRTLTLTY 852

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 2942292 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents.AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTCUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3990	93.6	801	1	US-08-725-012-2
2	1307.5	30.7	737	1	US-08-188-582-16
3	1307.5	30.7	737	1	US-08-646-715-16
4	775	18.2	921	1	US-08-188-582-2
5	775	18.2	921	1	US-08-646-715-2
6	237	5.6	2035	1	US-08-046-585-5
7	237	5.6	2035	1	US-08-393-703-5
8	237	5.6	2035	5	PCR-US93-11721-5
9	198	4.6	862	1	US-08-325-267A-4
10	197	4.6	2972	4	US-09-579-181-2
11	197	4.6	3118	4	US-09-579-181-1
12	191.5	4.5	894	3	US-08-362-525-22
13	191.5	4.5	894	3	US-08-971-692-15
14	184.5	4.3	1537	1	US-08-325-267A-2
15	178	4.2	1721	3	US-08-700-651-5
16	178	4.2	1721	3	US-08-928-361B-6
17	175.5	4.1	3969	4	US-08-061-376-5
18	172	4.0	750	4	US-09-165-239A-4
19	168	3.9	1837	3	US-08-928-361B-5
20	166	3.9	805	4	US-09-103-429A-4
21	164.5	3.9	752	1	US-08-244-189-2
22	164	3.8	903	2	US-08-853-310-2
23	160	3.8	1125	4	US-09-513-783A-152
24	160	3.8	1610	4	US-09-513-783A-22
25	158.5	3.7	786	4	US-09-103-429A-3
26	157.5	3.7	941	4	US-07-757-022B-14
27	157.5	3.7	1022	4	US-07-757-022B-84

28 157.5 3.7 1038 4 US-07-757-022B-74 Sequence 74, Appl
29 157.5 3.7 1049 4 US-07-757-022B-58 Sequence 58, Appl
30 157.5 3.7 1140 4 US-07-757-022B-104 Sequence 104, App
31 157.5 3.7 1270 4 US-07-757-022B-44 Sequence 44, Appl
32 157.5 3.7 1311 4 US-07-757-022B-42 Sequence 42, Appl
33 157.5 3.7 1313 4 US-07-757-022B-142 Sequence 142, Appl
34 157.5 3.7 1314 4 US-07-757-022B-50 Sequence 50, Appl
35 157.5 3.7 1320 4 US-07-757-022B-46 Sequence 46, Appl
36 157.5 3.7 1320 4 US-07-757-022B-60 Sequence 60, Appl
37 157.5 3.7 1354 4 US-07-757-022B-48 Sequence 48, Appl
38 157.5 3.7 1361 4 US-07-757-022B-40 Sequence 40, Appl
39 157.5 3.7 1363 4 US-07-757-022B-52 Sequence 52, Appl
40 157.5 3.7 1404 4 US-07-757-022B-2 Sequence 2, Appl
41 157.5 3.7 1404 4 US-07-757-022B-62 Sequence 62, Appl
42 157 3.7 2843 1 US-07-741-940-2 Sequence 2, Appl
43 157 3.7 2843 1 US-08-289-548A-2 Sequence 2, Appl
44 157 3.7 2843 1 US-08-452-654-2 Sequence 2, Appl
45 157 3.7 2843 2 US-08-370-235A-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-08-725-012-2

; Sequence 2, Application US/08725012

; Patent No. 5710025

; GENERAL INFORMATION:

; APPLICANT: Dikstein, Rivka

; APPLICANT: Tili, Robert

; TITLE OF INVENTION: B-Cell Specific Transcription Factor

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP

; STREET: 268 BUSH STREET, SUITE 3200

; CITY: SAN FRANCISCO

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION NUMBER:

; APPLICATION NUMBER: US/08/725,012

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: OSMAN, RICHARD A

; REGISTRATION NUMBER: 36,627

; REFERENCE/DOCKET NUMBER: B97-005

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 343-4341

; TELEFAX: (415) 343-4342

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 801 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-725-012-2

Query Match 93.6%; Score 3990; DB 1; Length 801;

Best Local Similarity 99.9%; Pred. No. 9,6e-319;

Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTLVTKVAPVSAPPKVSSGPRLPAPQIVAVKAPNTTTIQFPANLQLPPGVLIKNSGGL 60

Db 1 GTLVTKVAPVSAPPKVSSGPRLPAPQIVAVKAPNTTTIQFPANLQLPPGVLIKNSGGL 60

QY 61 MLVSPQOTVTRATTSNITSRPAPNPOTVKTCTVNSSQLKKVAVTPVKLAQIGT 120

Db 61 MLVSPQQTVAETTSNTTSRAVPANPQTVKICTVPNSSQLIKKVAIVTPVKLAQIGT 120
Qy 121 TVTTVPKPSVQSVAVPTSVVTPGKPLNTVTTLKPSILGASSTPSNEPLKAENSA 180
Db 121 TVTTVPKPSVQSVAVPTSVVTPGKPLNTVTTLKPSILGASSTPSNEPLKAENSA 180
Qy 191 VOINLSPMLNKKVCKNFKLWLKLAGSGSPGEMGNVKKLVEQLDADKIEAEFTPK 240
Db 181 VOINLSPMLNKKVCKNFKLWLKLAGSGSGSPGEMGNVKKLVEQLDADKIEAEFTPK 240
Qy 241 LYVELKSSPQHLVPLFKSVVALLROLPLNSQSFQOCVQOQSSDMVIATCTTIVTSPV 300
Db 241 LYVELKSSPQHLVPLFKSVVALLROLPLNSQSFQOCVQOQSSDMVIATCTTIVTSPV 300
Qy 301 VTTVSSSQSEKSIIVSGATAPRTVSVQTLNPLAGPVGAKAGVTVLHSGVPTAATGGTTA 360
Db 301 VTTVSSSQSEKSIIVSGATAPRTVSVQTLNPLAGPVGAKAGVTVLHSGVPTAATGGTTA 360
Qy 361 GTGLLQTSKPLVTSVANTVTVSLOPEKPVVSGTAVTLSLPAVTFGETSGAAICLPSVKP 420
Db 361 GTGLLQTSKPLVTSVANTVTVSLOPEKPVVSGTAVTLSLPAVTFGETSGAAICLPSVKP 420
Qy 421 VVSCWDHICKPVICTPVOIKLAQPGPVLSQAGIPTGSSSKQFSLFHVYVQPSGGNEK 480
Db 421 VVSCWDHICKPVICTPVOIKLAQPGPVLSQAGIPTGSSSKQFSLFHVYVQPSGGNEK 480
Qy 481 QVTTISHSTLTIOKCGQKTMVNTIIPTSQFPASILKQITLPCNKILSLQASPTQKNR 540
Db 481 QVTTISHSTLTIOKCGQKTMVNTIIPTSQFPASILKQITLPCNKILSLQASPTQKNR 540
Qy 541 IKENVTSQFREDDDINDVTSMAGVNLNENACILATNSLVTGLIQSCKDEFFLFIGALQ 600
Db 541 IKENVTSQFREDDDINDVTSMAGVNLNENACILATNSLVTGLIQSCKDEFFLFIGALQ 600
Qy 601 KRILDIGKKHDTIENSDAVNLISOATQERLGLLEKLTATAQHRMTYKASENYILCSD 660
Db 601 KRILDIGKKHDTIENSDAVNLISOATQERLGLLEKLTATAQHRMTYKASENYILCSD 660
Qy 661 TRSKLFLEKLDQLEKQKRDLEEREMLLKAASRSNKEDPQLRLKQAKELQLELAQI 720
Db 661 TRSKLFLEKLDQLEKQKRDLEEREMLLKAASRSNKEDPQLRLKQAKELQLELAQI 720
Qy 721 QHRDANLTATAAIGPKRKRPLESGIEGLKDNLLASGTSSTLTATQOLHRPRITRCLRLDI 780
Db 721 QHRDANLTATAAIGPKRKRPLESGIEGLKDNLLASGTSSTLTATQOLHRPRITRCLRLDI 780
Qy 781 FCMQEREMKYSRALLYALLK 801
Db 781 FCMQEREMKYSRALLYALLK 801

RESULT 2
us-08-188-582-16
; Sequence 16, Application US/08188582
; Patent No. 5534410
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Slegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA

ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,582
FILING DATE: 28-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 737 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-188-582-16
Query Match 30.7%; Score 1307.5; DB 1; Length 737;
Best Local Similarity 40.0%; Pred. No. 1.3e-98;
Matches 330; Conservative 103; Mismatches 200; Indels 193; Gaps 22;
Qy 13 PKPVSSG-----PRLPAQIVAVKAPNTTITIOFPAQLQPLPGTTLIKNSGGLM 61
Db 68 PTATTSIGIRATLTPVLAQLPQP-----PQNTPIQ---NFQLPPGMVLRSENGQLL 118
Qy 62 LVSPQQTVTR-----AETTSNITSRPAVPANPQTVKICTVPNSSQLIKKVAIVTPVKKL 115
Db 119 MI-PQALQAQQAQAHAQPTTMAPRPATPTSPAPVQISTVQAPGTPIIAR-QVTP---- 172
Qy 116 AQIGTVTTVPKPSVQSVAVPTSVVTPGKPLNTVT--TLKPSSILGASG-----TPS 168
Db 173 ---TTIIKQV---SQQTTPQPSATLQSRPGVQVQLVGGAAQTASLTATAVGTGTPQ 225
Qy 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFKLWLKLAGSGSGSPGEMGNVKKLVEQL 227
Db 226 RTVPGATTTSSAATE-----TMENVKCKNFKLWLKLAGSGKOSTETAAANVKELVQL 279
Qy 228 LDKATEAEFTTKLYVELKSSPQHLVPLFKSVVALLROLPLNSQSFQOCVQOQ-----TS 283
Db 280 LDGKIEAEFTSRLLYRELNSSPQPYLVPLFKRSLPALRLQTPDLSAAFTQQSQOQPPPPPTS 339
Qy 284 SDMVIATCTTIVTSPVTVTVSSSQSEKSIIVSGATAPRTVSVQTLNPLAGPVGAKAGV 343
Db 340 Q-----ATTALTAVLSSSVQRTAGTAATVTSALQPPVLSL----- 376
Qy 344 VTLHSGVPTAATGGTTAGTGLLQTSKPLVTSVANTVTVSLOPEKPVVSGTAVTSLPAV 403
Db 377 -----TQPTQVGVGKQGPPLVIQ-----Qppkp-----CALLRPQV 410
Qy 404 TFGETSGAAICLPSVKPVVSCWDHICKPVICTPVOIKLAQPGPVLSQAGIPTGSSSKQ 463
Db 411 TLTOT-----PMVALRQPH-NRIMLTTPQOIQL----- 437
Qy 464 LFLSFHVYVQPSGGNEKQVTTISHSSTLTIOKCGQKTMVNTIIPTSOFFPASPILKQITL 523
Db 438 ----- 451
Qy 524 PGNKIL-----SIQASPTQKNRIKENVTSQFREDDDINDVTSMAGVNLNENACILATNSL 580
Db 452 PGTKALSAVSAQAAAAQKKNLKEPGSGFRDDDDINDVASMAGVNLSEESARILATNSL 511
Qy 581 VGTLLQSCDKDEFFLFIGALQKRLDIDIGKKHDTIENSDAVNLISOATQERLGLLEKLT 640
Db 512 VGTLTRCKDETFLLQAPLQRRILEIGKKHGITELHPDQVSVYSHATQORLQNLVEKISE 571

QY 641 IAQHRMTTKYKASNYILCSQKFLKLEKLDQLEKQKRDLEERMLLKAASRSNKEDP 700
DB 572 TAQKNFYSKDDRYEQASDVRAQLKFFELQDQIEKQKDEQEREILMRAAKSRSDP 631
QY 701 EQLRLKQAKELQOQLAQIHRDANLATAAIGPRKKRPLE-----SGIEGLKDNLLAS 755
DB 632 EQLRLKQAKEMQOQELAQMRQDANLATAAIGPRKKRVDCPGPGSGAEGSGPGSVVP 691
QY 756 GTSSTATKQLHRPRITRCLRDILFCMEQEREMKYSRALLYLALK 801
DB 692 GSSGVGTPTQRTQRTITRNLRLIFCLENERETSHSLLYKAFK 737

RESULT 3

US-08-646-715-16
; Sequence 16, Application US/08646715
; Patent No. 5637686
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,715
; FILING DATE: 09-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/188,582
; FILING DATE: 28-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 737 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-646-715-16

Query Match 30.7%; Score 1307.5; DB 1; Length 737;
Best Local Similarity 40.0%; Pred. No. 1.3e-98;
Matches 330; Conservative 103; Mismatches 200; Indels -193; Gaps 22;

QY 13 PKVSSG-----PRLPAQIVAVKAPNTTITQFPANLQPLPGVLIKNSGGLM 61
DB 68 PTATTSIGIRATLPTVLAFLRLPQP-----PQNPTNIQ---NFQLPPGMVLRSENGQL 118

QY 62 LVSPQQTVTR-----AETTSNITSRAVPANPQTVKICTVPSNSSOLIKKAVTPVKKL 115
DB 119 MI-POQALQAOQAHAQPTTMAPRAPTSPAPPVISTVAPGPPIIAR-QVTP----- 172
QY 116 AQIGTAVTVPKPSVQSVAVPTSVVTPGKPLNTVT--TLKPSSSLGASS-----TPS 168
DB 173 ----TWIIKQV---SOAQITVQPSAYLQRSFGVQPOLYLCGAATASLGATATVQTGTPQ 225
QY 169 NE-PNLKAENSAAVQINLSPTMLENKYKKNFLAMLIKACSGSQSPSEMGNKKVLEQL 227
DB 226 RTVPGATTSSAATE-----TMENVKCKNFKLSTLIKASSGKQSTETAAANKVELYONL 279
QY 228 LDKAIEAEETRKLYVELKSSQPHLVPLKSKVVALRQLLPNSQSFQOCVQ-----TS 283
DB 280 LDGKIEAEDFTSRLYRELNSSPQYLVFUKRSLPALRLTPDSAAFIQOQQPPPTPS 339
QY 284 SDMVIATCTTVTTPVTSSSQSEKIIIVSGATAPRTSVQTLNLAGPVGAKGV 343
DB 340 Q-----ATTALTAVVLSSVQRTAGKTAATVTSALQPPVLSL----- 376
QY 344 VTLHVGPTAATGGTAGTGLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAVTSLPAV 403
DB 377 -----TQPTQVGKQGOPTPLVIQ-----QPKP-----GALIRPPQV 410
QY 404 TFCETSGAAICLPSVKPVVSCWDHICKPVIGTPVQIKLQAPGVLSQAPAGIPTGSSSKQ 463
DB 411 TLTQT-----PMVALRQPH-NRIMLTTPQIQ----- 437
QY 464 LFLSFHVQVQPSGGNEKQVTTISHSSTLTIQCGOKTTPVNTIPTTSQFPPASTLKQITL 523
DB 438 -----NPLQPVVVKRAVL 451
QY 524 PGNKIL---SLQASPTQKNRIKENVTSCFDEDDINDVTSMAGVNLNEENACILATNSEL 580
DB 452 PGTKALSASVSAQAAAKKKEPGGSRDDDDINDVASWAGVNLSEESARILATNSEL 511
QY 581 VGTLIOSCKDEPFLFCALQKRIIDIGKKHIDITELNSDANLISOATQERLGLLEKLA 640
DB 512 VGTLTRSCKDETFELQAPLQRRILEIGKKGITELHPDVVSYVSHATQOQLNLEKISE 571
QY 641 IAQHRMTTKYKASNYILCSQKFLKLEKLDQLEKQKRDLEERMLLKAASRSNKEDP 700
DB 572 TAQKNFYSKDDRYEQASDVRAQLKFFELQDQIEKQKDEQEREILMRAAKSRSDP 631
QY 701 EQLRLKQAKELQOQLAQIHRDANLATAAIGPRKKRPLE-----SGIEGLKDNLLAS 755
DB 632 EQLRLKQAKEMQOQELAQMRQDANLATAAIGPRKKRVDCPGPGSGAEGSGPGSVVP 691
QY 756 GTSSTATKQLHRPRITRCLRDILFCMEQEREMKYSRALLYLALK 801
DB 692 GSSGVGTPTQRTQRTITRNLRLIFCLENERETSHSLLYKAFK 737

RESULT 4

US-08-188-582-2
; Sequence 2, Application US/08188582
; Patent No. 5534410
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco

```
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,582
FILING DATE: 28-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-188-582-2

Query Match      18.2%   Score 775;   DB 1;   Length 921;
Best Local Similarity 29.6%   Pred. No. 8.8e-55;
Matches 258; Conservative 128; Mismatches 308; Indels 178; Gaps 26;

Qy 25 PQIVAVKAPNTTITQFPANLQPPGTVLKNSGFLML-----VSPQOTVTRAETTS 76
Db 133 PQSPSITLTLNTGTGA-----LLVKTGDFQLLRVGTGTPPTVTQTIINTSNNS 184
Qy 77 NITSRPVAPNPQVKTICITVPSNSSQ-----LIKKVATPVKKLAQIGTIVVTVP 127
Db 185 NITTSNTHPTTQ-IRLQTPAAASMTNTATSNIVNSVASSGVNSQPPHLTQLNAQ 243
Qy 128 KPSSVQSVAVPTSVVTPVTPKPLNTVTLKPSSILGASGTPSNPNLKAENSAAVQINLSP 187
Db 244 APQLPQITQITIPAAQSQQQQVNVSSAGGTATAVSSTA-----AT 287
Qy 188 TMLNVK-KCKNFLAMLIKACSGSPGEMGONVKVLEQLDLAKIAEEFTRKLYVELK 246
Db 288 TQQNTKEKCKFLANLIEL--STREPKVKEKVTNLIQELVNAVNEPEECDRLERLN 345
Qy 247 SSPQHLVFLKSKVVALRQL-----LPSNQSFQIQ--- 276
Db 346 ASPQCLIGFLKSLPLRLRQALYTKELVIEGKPPQHVGLAGLSQQLPKIQAIPIG 405
Qy 277 ---OCVOOTSSDMVIATCTTTVTTSPTVTVTVSSQSEKSIIVSGATAPRTVS---VQT 329
Db 406 PSQTTIGQTVRMIL--TPNALGTPTPTGHTTISKQPN---IRLPTAPRLVNGGIRT 460
Qy 330 LNPAGVPKAGVWVTHLSVGTAAATGTTAGTGLLQTSKPLVTSVANTVTVTSIQPEKP 389
Db 461 QIP-SLQVPGQANIVQIR--CPQHAQLQRTGSGVQIRATRP-----PNSVPTAN----- 506
Qy 390 VVSGTAVTSLPANTFGTSGAAICLPSKVVPVSEFWDHCKPVIGTPVQIKLAQPGVUL 449
Db 507 -----KLTAVKVGQTIKAI--TPSLHP-----PSLAAISGGP-----PPTPTL 543
Qy 450 SOPAGIPTGSSSK---QLFSLFHVQQPSGGNEKQVTTISHSS----- 489
Db 544 SVLSTLSASATITLPIPSLPTVHLPPPEALRAREQNSLNHNSHFDKALVEIKAPSLHP 603
Qy 490 -----TLTIQKCGKMPVNTIIPTSQFPASPILKQITLPGNKILSIQ-----AS 534
Db 604 PHMERINASLTPIGAKTM-----ARPPPAINKAIGKKKRDAMENDAKLNTSSGGA 654
Qy 535 PTQKNRIKENVTSFRDEDDINDVTSAGVNLNEENACILATNSLVELTTLQSCCKDEPFL 594

STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,582
FILING DATE: 28-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-188-582-2

Query Match      18.2%   Score 775;   DB 1;   Length 921;
Best Local Similarity 29.6%   Pred. No. 8.8e-55;
Matches 258; Conservative 128; Mismatches 308; Indels 178; Gaps 26;

Qy 25 PQIVAVKAPNTTITQFPANLQPPGTVLKNSGFLML-----VSPQOTVTRAETTS 76
Db 133 PQSPSITLTLNTGTGA-----LLVKTGDFQLLRVGTGTPPTVTQTIINTSNNS 184
Qy 77 NITSRPVAPNPQVKTICITVPSNSSQ-----LIKKVATPVKKLAQIGTIVVTVP 127
Db 185 NITTSNTHPTTQ-IRLQTPAAASMTNTATSNIVNSVASSGVNSQPPHLTQLNAQ 243
Qy 128 KPSSVQSVAVPTSVVTPVTPKPLNTVTLKPSSILGASGTPSNPNLKAENSAAVQINLSP 187
Db 244 APQLPQITQITIPAAQSQQQQVNVSSAGGTATAVSSTA-----AT 287
Qy 188 TMLNVK-KCKNFLAMLIKACSGSPGEMGONVKVLEQLDLAKIAEEFTRKLYVELK 246
Db 288 TQQNTKEKCKFLANLIEL--STREPKVKEKVTNLIQELVNAVNEPEECDRLERLN 345
Qy 247 SSPQHLVFLKSKVVALRQL-----LPSNQSFQIQ--- 276
Db 346 ASPQCLIGFLKSLPLRLRQALYTKELVIEGKPPQHVGLAGLSQQLPKIQAIPIG 405
Qy 277 ---OCVOOTSSDMVIATCTTTVTTSPTVTVTVSSQSEKSIIVSGATAPRTVS---VQT 329
Db 406 PSQTTIGQTVRMIL--TPNALGTPTPTGHTTISKQPN---IRLPTAPRLVNGGIRT 460
Qy 330 LNPAGVPKAGVWVTHLSVGTAAATGTTAGTGLLQTSKPLVTSVANTVTVTSIQPEKP 389
Db 461 QIP-SLQVPGQANIVQIR--CPQHAQLQRTGSGVQIRATRP-----PNSVPTAN----- 506
Qy 390 VVSGTAVTSLPANTFGTSGAAICLPSKVVPVSEFWDHCKPVIGTPVQIKLAQPGVUL 449
Db 507 -----KLTAVKVGQTIKAI--TPSLHP-----PSLAAISGGP-----PPTPTL 543
Qy 450 SOPAGIPTGSSSK---QLFSLFHVQQPSGGNEKQVTTISHSS----- 489
Db 544 SVLSTLSASATITLPIPSLPTVHLPPPEALRAREQNSLNHNSHFDKALVEIKAPSLHP 603
Qy 490 -----TLTIQKCGKMPVNTIIPTSQFPASPILKQITLPGNKILSIQ-----AS 534
Db 604 PHMERINASLTPIGAKTM-----ARPPPAINKAIGKKKRDAMENDAKLNTSSGGA 654
Qy 535 PTQKNRIKENVTSFRDEDDINDVTSAGVNLNEENACILATNSLVELTTLQSCCKDEPFL 594

US-08-646-715-2
Sequence 2, Application US/08646715
Patent No. 5637686
GENERAL INFORMATION:
APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dynlacht, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAPS AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESS: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,715
FILING DATE: 09-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/188,582
FILING DATE: 28-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-646-715-2

Query Match      18.2%   Score 775;   DB 1;   Length 921;
Best Local Similarity 29.6%   Pred. No. 8.8e-55;
Matches 258; Conservative 128; Mismatches 308; Indels 178; Gaps 26;

Qy 25 PQIVAVKAPNTTITQFPANLQPPGTVLKNSGFLML-----VSPQOTVTRAETTS 76
Db 133 PQSPSITLTLNTGTGA-----LLVKTGDFQLLRVGTGTPPTVTQTIINTSNNS 184
Qy 77 NITSRPVAPNPQVKTICITVPSNSSQ-----LIKKVATPVKKLAQIGTIVVTVP 127
Db 185 NITTSNTHPTTQ-IRLQTPAAASMTNTATSNIVNSVASSGVNSQPPHLTQLNAQ 243
Qy 128 KPSSVQSVAVPTSVVTPVTPKPLNTVTLKPSSILGASGTPSNPNLKAENSAAVQINLSP 187
Db 244 APQLPQITQITIPAAQSQQQQVNVSSAGGTATAVSSTA-----AT 287
Qy 188 TMLNVK-KCKNFLAMLIKACSGSPGEMGONVKVLEQLDLAKIAEEFTRKLYVELK 246
Db 288 TQQNTKEKCKFLANLIEL--STREPKVKEKVTNLIQELVNAVNEPEECDRLERLN 345
Qy 247 SSPQHLVFLKSKVVALRQL-----LPSNQSFQIQ--- 276
Db 346 ASPQCLIGFLKSLPLRLRQALYTKELVIEGKPPQHVGLAGLSQQLPKIQAIPIG 405
Qy 277 ---OCVOOTSSDMVIATCTTTVTTSPTVTVTVSSQSEKSIIVSGATAPRTVS---VQT 329
Db 406 PSQTTIGQTVRMIL--TPNALGTPTPTGHTTISKQPN---IRLPTAPRLVNGGIRT 460
Qy 330 LNPAGVPKAGVWVTHLSVGTAAATGTTAGTGLLQTSKPLVTSVANTVTVTSIQPEKP 389
Db 461 QIP-SLQVPGQANIVQIR--CPQHAQLQRTGSGVQIRATRP-----PNSVPTAN----- 506
Qy 390 VVSGTAVTSLPANTFGTSGAAICLPSKVVPVSEFWDHCKPVIGTPVQIKLAQPGVUL 449
Db 507 -----KLTAVKVGQTIKAI--TPSLHP-----PSLAAISGGP-----PPTPTL 543
Qy 450 SOPAGIPTGSSSK---QLFSLFHVQQPSGGNEKQVTTISHSS----- 489
Db 544 SVLSTLSASATITLPIPSLPTVHLPPPEALRAREQNSLNHNSHFDKALVEIKAPSLHP 603
Qy 490 -----TLTIQKCGKMPVNTIIPTSQFPASPILKQITLPGNKILSIQ-----AS 534
Db 604 PHMERINASLTPIGAKTM-----ARPPPAINKAIGKKKRDAMENDAKLNTSSGGA 654
Qy 535 PTQKNRIKENVTSFRDEDDINDVTSAGVNLNEENACILATNSLVELTTLQSCCKDEPFL 594
```

Matches 256; Conservative 128; Mismatches 308; Indels 178; Gaps 26;

QY 25 PQIVAVKAPNTTIOFPANLQLPCTVLIKNSGPMML-----VSPQOTVTRAETTS 76
Db 133 POSPISITLSTLNTGTGA-----LLVKTDFGQLLRVGTGTGPTVTQITNTSNN 184
QY 77 NITSRPVAPNPQVVKICTVPNSSQ-----LIKKAIVTPVKKLAQIGTIVTTP 127
Db 185 NTTSTNHPHTTQ-IRLOTVPAASMTNTATSLIIVNSVASSGYANSSQPHLTQLNAQ 243
QY 128 KPSSVQSVAVPTSVVTPGKPLNVTTLKPSLSGASTPSNEPNLKAENSAVAINLSP 187
Db 244 APQLPQITQITIPAQSOQOVNNSVAGGTATAVSSTA-----ATT 287
QY 188 TMLNVR-KCNFLAMLIKACSSQSPENCONVKLVQLLDAKIAEETRLKLYELK 246
Db 288 TOQNTKEKCKRLANLIEL--STREPKVKEKNVRTLQIELVNAVPEEFCDLRELLN 345
QY 247 SSPQHLVPLFLKKSVALRQL-----LPNSQSFIO--- 276
Db 346 ASPQCLIGFLKKSPLLRQALYTKELVIEIKPPQHVGLGAGLSQQLPQIAQIRPIG 405
QY 277 ----OCVOOTSSDMVIACTTFTVTTPSVVTTTVSSSSEKSIIVSGATAPRTVS---VQT 329
Db 406 PSQTTTIGQTOVRMI--TPNALGTPTPTIGHTTISKOPPN---IRLPTAPRLVNTGGIRT 460
QY 330 LNPLAGVPKAGVVTLSHVGPATAAGTTAGTGLLOTSRPLVTSVANTVTVLSLOPEKP 389
Db 461 QIP-SLOVPGQANIVQIR--GPQHAQLORTGSVQIRATRP-----PNSVPTAN----- 506
QY 390 VVSGTAVTSLPAPVFTGETSGAAICLPSKPVVSWFCWDHICKPVIGTPVQIKLAQPGPVL 449
Db 507 -----KLTAVKVQGTQIKAI-TPSLHP-----PSLAAISGGP-----PPTPTL 543
QY 450 SOPAGITPGSSK---OLFSLFHVQVQPSGGNEKQVTTISHS----- 489
Db 544 SVLSTLNSASTTLPISLPTVHLPPALRAEQMNSLNHNSHFDKALVEIKAPSLHP 603
QY 490 -----TLTIQCKCKMPVNTIIPTSQFPASPILKQITLPGNKILSLQ-----AS 534
Db 604 PHMERINASLTPICAKTM-----ARPPPAINKAIGKKRDRAMEMDAKLNTSSGAA 654
QY 535 PTKNRKENVTSFREDDOINDVTSMAGVNLNEENACILATNSLSELVGTLIQSKDEPFL 594
Db 655 SAANSFQOQSSMSMYGDDNDVAMGYNLAESQRIKCC-TENIGTQIRSKDEVEL 713
QY 595 FIGALQKRIIDIGKHHIDITELNSDAVNLISOATOERLGLLEKLTATAOHRMTYTKASEN 654
Db 714 NLPSLQARIRAITSEAGLDEPSQDVAVLISHACQERLKNIVEKLAVIAEHRIDVILKDP 773
QY 655 YILCSDTRSQLKFLKLDQLEKQKRDLEEREMLLKAASRKNKEDPQLRLKQKAKELQ 714
Db 774 YEPADYVGOIKFLEELDQAKQKHELEERMLLRAAKSRVDEPQAKMKARAKEMOR 833
QY 715 LELAQIHRDANLTATAAIGPKRRKPLE-----SGIEGLKDNLLASGTSSTLATAKQLHRP 769
Db 834 AEMELQRDANLTALQAIGPKRKLKLDGETVSSGASGGCVLSSSGSAPPTL-----RP 889
QY 770 RITRICLRDLIFCMEQREMKYSRALLYLALIK 801
Db 890 RIKRVNLRDLIFYMEQEREFRCRSMLEKTYLK 921

RESULT 6
US-08-046-585-5
; Sequence 5, Application us/08046585
; Patent No. 5453362
; GENERAL INFORMATION:
; APPLICANT: Lamarco, Kelly
; APPLICANT: Wilson, Angus
; APPLICANT: Herr, Winship
; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
; TITLE OF INVENTION: HOST CELL FACTOR

NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embardadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/046,585
APPLICATION NUMBER: US/08/046,585
FILING DATE: 12-APR-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A.
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57503-1/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-046-585-5

Query Match 5.6%; Score 237; DB 1; Length 2035;
Best Local Similarity 22.6%; Pred. No. 4.2e-10;
Matches 148; Conservative 84; Mismatches 239; Indels 184; Gaps 32;

QY 10 VSAPPKVSSGPRLPAPQIVAVKAPNTTIIQPPANLQLPKPPVLIKNSGPMMLVSPQOTV 69
Db 537 IGSSPSSGMAALAAAAAATOKIPSSA---PTVLSVPAGTTIVKT-----MAVTPGTTT 588
QY 70 TRAEITSNITSRRPAPNPOTVKTCTVPNSSSOLIKKVVAVTPVKKLAQIGTIV----- 122
Db 589 LPA--TVKASSPVVWSNPAT-----RMLKTA-----AQVTSVSSATNTS 628
QY 123 ---VTTVPKPSVQSVAVPTSVVTPGKPLNTVTLK-PSSL-GASSTPSNEPNL----- 173
Db 629 TRPIITVHKSGTV-TVAQQAQVTVTVGGVTKTITLVKSPISVPGGSAIINLKVMSV 687
QY 174 --KAENSAVAINLSPMTLENVKKCKNPL--AMLIKACSSQSP-----EMGQNVK 221
Db 688 QTKPVQTSVAVTQASTGPTVQIQTKGPLPAGTILKLVTSADGKPTTIIITTTQASGACTK 747
QY 222 KLVQLLDAKIAEETRLKLVKSSPQHLVPLFKSVVALLQPLNPSQSFIOQVQ 281
Db 748 PTILGI-----SSVSPSTT---KPTTTIITKTPMSAITIQAAGATG 785
QY 282 TSSDMVIATCTTIVTT-----SPVVT-----TVSSSQSEKSIIVSGA-----TA 321
Db 786 VTSSPGIKSPITIIITTKVMTSGTGAPAKIITAVPKIATGCGQGVTVQVVLKAGAPQGTI 845
QY 322 PRT-----VSVOTLNPAGPVCAK--AGVTLHVSVPFAATAGTTTAGTGLLOTS 368
Db 846 LRTVPMGGVRLVTPVTVSAVKPAVTTLVVKGTTGTTVLGTGTGVST--SLAGAGGHSTS 903
QY 369 KPLVTSVA--NVTVTVLSLOPEKPVVSGTAVTSLPVTFTGETSGAAICLPSVKPVVSWFCW 426
Db 904 ASLATPITTLTGTTIATLSSQ-----VINPITAVTSAQVTLTAAGGLTPTITMQPV----- 954
QY 427 DHICKPVIGTPVQIKL-AQCPGVLSQPA-GIPTGSSSKQLFLSWFWQVQPSGGNEKQVTT 484
Db 955 -----SQPTQVTLITAPSGVEAQPVDLPLVS-----ILASP-----TT 987

QY 485 ISHSSTLTQKCGQKMPVNTIPTISQFPASI-----LKQITLPGNKILSLQASPTQ 537
Db 988 EQPTATVTIADSGQGVQGTVTLCVSNPPCETHETGTTNTATTIVAN--LGGHPQPTQ 1045
QY 538 KNRIVENVTSQFDEDDINDVTSMAGVNLEENACILATNSLVLGTLIOCKDEP 592
Db 1046 VQFV-----CDRQEAASLVTSTVG-----QQN-----GSVVRVCSNPP 1079

RESULT 7
US-08-393-703-5
; Sequence 5, Application US/08393703
; Patent No. 5585239
; GENERAL INFORMATION:
; APPLICANT: Lamarco, Kelly
; APPLICANT: Wilson, Angus
; APPLICANT: Herr, Winship
; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
; TITLE OF INVENTION: HOST CELL FACTOR
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/393,703
; FILING DATE: 24-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57503-2/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-393-703-5

Query Match 5.6%; Score 237; DB 1; Length 2035;
Best Local Similarity 22.6%; Pred. No. 4.2e-10;
Matches 148; Conservative 84; Mismatches 239; Indels 184; Gaps 32;

QY 10 VSAPPKVSGLPLAPQIVAVKAPNTTTTQFPANLQLPPTGVLIKNSGPLMLVSPQQT 69
Db 537 ICSSQMSGMAALAAAAATQIPSSA---PTVLSVPAGTTIVKT-----NAVTPGTTT 588
QY 70 TRAEITSNTSRPVPANPQTKVICTVPNSSSOLIKKAVTPVKKLAQIGTTV-----122
Db 589 LPA--TVKVASSPVMVSNPAT-----RMLKTA-----AQVTSVSSATNTS 628
QY 123 ---VTVPKPSSVQSVAVPTSVVTPGKPLNTVTLK--PSSL-GASSTPSNEPNL----173
Db 629 TRPIITVHKSGTV--TVAQQAQVTVTVGGVTKTITLVKSPISVPGSGSALISNLGKVMVW 687
QY 174 --KAENSAVQNLSPMLNENKCKNFL--AMLIKACSGSQSP-----EMQNVK 221
Db 688 QTKPVQTSVTCQAGTGPVTVIIQIKGKPLPAGTILKLVTSADCKPTTIITTTQASGAGTK 747

QY 222 KLVEQLLDAKIEAEERKLYVELKSSPOPHLVFPLKSVVALROLLPNSQFIQCVCQ 281
Db 748 PILGI-----SSVSPSYT---KPGTTTIKIPMSAITQAGATG 785
QY 282 TSDMVIATCTTTVT-----SPVVT-----TVSSSQSEKSIIVSGA-----TA 321
Db 786 VTSSPGIKSPITITTKVMTSGTGAPAKIITAVPKIATGHGQGVTVVVLKGAQGPQGTI 845
QY 322 PRT-----VSQTLNPLAGVPGAK--AGVTLHSVGPATAGTTAGTGLQTS 368
Db 846 LRTVPMGGVRLVTPVTSVAVKPAVTTLVKGTGTTLGTGTGTST--SLAGAGHSTS 903
QY 369 KPLVTSVA--NTVTVSLQPEKPVVSGTAVTSLPAVTFGETSGAAICLPSVKPVVSEFW 426
Db 904 ASLAPITTLGTIATLSSQ-----VINPTAITVSAAGTTTAAAGLLTPTITMPV-----954
QY 427 DHICKRVICTPVOIKL-AQGPVLSQPA-GIPTGSSSKQLFLSFHVVOQPSGSGNEKQVTT 484
Db 955 -----SQPTQVTLITAPSGVEAQPVDLPVS-----ILASP-----TT 987
QY 485 ISHSSTLTQKCGQKMPVNTIPTISQFPASI-----LKQITLPGNKILSLQASPTQ 537
Db 988 EQPTATVTIADSGQGVQGTVTLCVSNPPCETHETGTTNTATTIVAN--LGGHPQPTQ 1045
QY 538 KNRIVENVTSQFDEDDINDVTSMAGVNLEENACILATNSLVLGTLIOCKDEP 592
Db 1046 VQFV-----CDRQEAASLVTSTVG-----QQN-----GSVVRVCSNPP 1079

RESULT 8
PCT-US93-11721-5
; Sequence 5, Application PC/TUS9311721
; GENERAL INFORMATION:
; APPLICANT: Lamarco, Kelly
; APPLICANT: Wilson, Angus
; APPLICANT: Herr, Winship
; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
; TITLE OF INVENTION: HOST CELL FACTOR
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11721
; FILING DATE: 03-DEC-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: FP-57503-1/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US93-11721-5

```
Query Match      5.6%  Score 237; DB 5; Length 2035;
Best Local Similarity 22.6%  Pred. No. 4.2e-10;
Matches 148; Conservative 84; Mismatches 239; Indels 184; Gaps 32;

QY 10 VSAPPKSSGRLPAPQIVAVKAPNTTIIOPANLQPLPGTGLIKSNGLMLVSPQOQTV 69
Db 537 IGSSPQSGMAALAAAAATOKIPSSA---PTVLSVPAGTTIVKT-----NAVTPGTTT 588
QY 70 TRAEITSNITSRPAVPANPQIVKICTVPNSSSOLIKKVVAVTPVKLAQIGTIV-----122
Db 589 LPA---TVKVASSPVNVSNPAT-----RMLKTA-----AOVGTFSVSATNTS 628
QY 123 ---VTVPKPSSVOSVAVPTSVWTPGKPLNTVTTLK-PSSL-GASSTPSNEPNL-----173
Db 629 TRPIITHKSGTV-TVAQQAQVTVTVGGVTKTTLVKSPISVPFGGSAALNLGKVMVSV 687
QY 174 --KAENSAVQINISPTMLENVKCKNPL--AMLIKACSGSQSP-----EMGNQVK 221
Db 688 QTKPVQTSVAVTQOASTGPVTQIIQTKGPLPAGTILKLVTSADGKPTTIIITTTQASGAGTK 747
QY 222 KLVQLLDKAEAEFFRKLVKELKSSPOPHLVFPLKSVVALRQLLPNSOFTQOQVQ 281
Db 748 PTILGI-----SSVSPSTT---KPGTTTIITKIPMSAIIITQAGATG 785
QY 282 TSDMWIATCTTIVT-----SPVVT-----TVSSSQSEKSIIVSGA-----TA 321
Db 786 VTSFPGIKSPIITITTKVMTSGTGAPAKIITAVPKIATGCGOQVTVQVVLKAGAPQPGTI 845
QY 322 PRT-----VSVOTLNPLAGPVGAK--AGVTLHSGVPTAAGTTAGTGLLQTS 368
Db 846 LRTVPMGVRVLTVPVTSVAKPAVTTLVKGTGTGTTILGTGTGVST--SLAGAGHSTS 903
QY 369 KPLVTSVA--NTVTVLSQPKPVVSGTAVTSLPAPVTFGTSGAAICLPSKVPVSPCW 426
Db 904 ASLATPITTLGTIATLSQ-----VINETAIVSAQTTTLTAAGLTTTPTITMQPV-----954
QY 427 DHICKPVIGTVPQIKL-AQGPVLISOPA-GIPTSSSKQLFSLFHVQOOPSGGNEKQVTT 484
Db 955 -----SQTQVTLITAPSGVCAQVHDLPVS-----ILASP-----TT 987
QY 485 ISHSTLTIQCGQKTMPTNIIPTSPFPASI-----LKQITLPGNKILSLQASPTQ 537
Db 988 EQPATVIADSGQDVGQGVTVLVCNPPCPCETHETGTNTATTIVVAN--LGHPQPTQ 1045
QY 538 KNRKENVTSFRDEDDINDVTSMAGVNLNEENACILATNSELVGTLIQSKDEP 592
Db 1046 VQFV-----CDROEAAASLTSTVG-----QQN-----GSVVRVCNPP 1079

RESULT 9
US-08-325-267A-4
; Sequence 4, Application US/08325267A
; Patent No. 5585271
; GENERAL INFORMATION:
; APPLICANT: WATARI, JUNJI
; APPLICANT: TAKATA, YOSHIHIRO
; APPLICANT: OGAWA, MASAHIRO
; APPLICANT: PENFILLA, MERJA
; APPLICANT: ONNELA, MAIJA-LEENA
; APPLICANT: KERANEN, SIRKKA
; TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST
; CONTAINING THEM
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, MCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/325,267A
FILING DATE: 18-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP PCT/JP94/00290
FILING DATE: 24-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 38871/1993
FILING DATE: 26-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2589-023-0XPCT
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 862 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-325-267A-4

Query Match      4.6%  Score 198; DB 1; Length 862;
Best Local Similarity 19.1%  Pred. No. 1.7e-07;
Matches 145; Conservative 110; Mismatches 278; Indels 228; Gaps 29;

QY 11 SAPPKVSSGRLPAPQIVAVKAPNTTIIQF---PANLQPLPGTGLIKSNGL-----60
Db 199 SLPPNIEGTVMYAGYYPMKVYSNVSWCTLPISVLPDGTIVSDDFEGVYVSFDDDL 258
QY 61 -----MLVSPQOVTTRAET-----TSNITSRPAVPANPQIVKICTVPNS 99
Db 259 SOSNCTVPDPSNAYASTTTTTPWTGTFTSTSTMTTGTNGVPTD-ETVIVIRPTT 317
QY 100 SSOLIKKVVAVTPVKLAQIGTIVTVTPKSSVOSVAVPTSVWTPGKPLN-TVTTLK- 157
Db 318 AS-----TIIITTEPWTGFTTSTSTMTTGTNGQPTDETVIVIRT 359
QY 158 PSSLGASSTPSNEP-----NLKAENSAVQINISPTMLENVKCKNPLAMLIKACSGS 211
Db 360 PTEGLVTT-TTEPWTGFTTSTSTMTSTVTCNGLPTD-ETVIVVKT-PTTAISSSLSS 416
QY 212 QSPMGQNVKLVQLLDKAEAEFFRKLVKELKSSPOPHLVFPLKSVVALRQLLPNS 271
Db 417 SSGQITSSI-----TSSRPITTPFPS-----NG 440
QY 272 QSFQOQVQOQTSMDVIATCTTV-TTSPVVTTVSS-----SQSEKSIIV-----316
Db 441 TSVI-----SSSVISSVTSLSFTSPVSSSVSSSTTSIFSEKSSKSVIPTSS 493
QY 317 --SCATAPRTVSVQTLNPLAGPVGAKAGVTLHSGVPTAAGTTAGTGLLQSKPLVTS 374
Db 494 STSGSESESTSS-----AGSVSSSPFISSEKSKSPYSS-----SSLPLVTS 535
QY 375 VANTVTVLSQPEKPVVSGTAVTSLPAPVTFGTSGAAICLPSKVPVSVFCWDHIC-----430
Db 536 ATTSQETAS-----SLPATTITTEQTTLV-----TVTSCESHVCTESI 575
QY 431 KPVITGPVQIKLAQPGPVLSPQAGIPTGSSSKQLFSLFHVQOQPSGGNEKQVTTISHST 490
Db 576 SPAIVSTATVTV-----SGVTTEYTWCPISITETTKQKGTETQTTKQTV 625
QY 491 LTIQK-----CQKTMPTNII-----IPTSQFPPASILKQITLPGNK 527
Db 626 VTISCESDVCSKTASPAIVSTSTATINGVTTEYTWCPISITETTSRQOQTLVTVTSCSG 685
QY 528 ILSQASP-----TQKNRIKENVT-----SCFRDEDDINDVTSMAGVNLNEENACILAT 576
```

Db 686 VCSETAIPVATATVNDVTVYTPWRPQTANEESVSKMNSATGETTTNTLAAETTT 745
QY 577 NSELVGTLIQSCKDEPFLFIGALQKR-----ILDIGKKHDITELN 616
Db 746 NTVAAETIITNGAETKTVVTSLSRSHAEQTASATDVIGHSSSVSVSETGNTKSLT 805
QY 617 SDAVNLISQATRLGLLEKLTATIAHQHRTWYKASENYIL 657
Db 806 SSGLSMTSQOPRSTPASSVMGVST-ASLEISTYAGSANSLL 845
RESULT 10
US-09-579-181-2
; Sequence 2, Application US/09579181
; Patent No. 6365372
; GENERAL INFORMATION:
; APPLICANT: Chrivia, John
; APPLICANT: Yaciuk, Peter
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247
; CURRENT APPLICATION NUMBER: US/09/579,181
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,620
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2972
; TYPE: PRT
; ORGANISM: Human
US-09-579-181-2

Query Match 4.6%; Score 197; DB 4; Length 2972;
Best Local Similarity 23.7%; Pred. No. 1.5e-06;
Matches 140; Conservative 75; Mismatches 240; Indels 136; Gaps 32;
QY 2 TLVTKVAPVSAPPKVSQG-----PRLPAPQIVAVKAPNTTTIQFPANLQL-- 46
Db 1063 TPVPLAPAPRPP--SSGLPAVLNPRPTLTGRLPTTLGTARAPMPTTLVRPLKLKLVH 1120
QY 47 --PPGTVLKNSGGLMLVSPQQTVTRAETTSNITSRPAVPANPQVVKICTVPSNSSQLI 104
Db 1121 SPSEFVSASAGNAPLTISPLHVP-----SSLPQSPASPM-----PIPNSSPLAS 1166
QY 105 KKVAVTPVKKLAQIGTVTVTVTPKPSVQSVAVPTSV-VTVTPGKP--LNTVT-TLKPS 160
Db 1167 PVSSTVSFVPLSSSLPISVPTTLAPASA-PLTIPISAPLTVSASGALLTSVTPPLAPV 1225
QY 161 LGASSTPSNEPNLKAENSAVQINL--SPTMLENVKCKNFAMLIKLCACSGSQSPMGQ 218
Db 1226 PAAPGPPLQPSGASPSALTLGLATAPSLSSQTPGHPLL-----LAPTSSHVPGLNS 1280
QY 219 NVKKLVEQ-LLDAKIEAEFEFRKLYVELKSSQPH-----LVPFLKKSVMALRQLLPS 271
Db 1281 TVAPACSPVLVPASALASPF-----PSAPNPAQAASLLAPASSASOALATPLAPMA 1332
QY 272 QSFIOQCVQQTSDMVIATCTTTVTTSVVTTTVSSSQSEKSIIVSGAT----- 320
Db 1333 AP--QTAILAPSPAPPLAP-LPVLAPSPGAAPVLASSQTPVPVMAFSPSTPGTSLASAPV 1389
QY 321 -APRTV-----SVQTL--NPLAGPVGAKAGVTVTLHVSQGTAA--TGCT-----TAGTG 365
Db 1390 PAPTVPVLAPSSQTMLPAPVPSPLPSPASTQTL-ALAPALAPTLGGSSPSQTLSLGTGNP 1448
QY 366 QTSKPLVTVANTVTVSVLSQEPKPVVSGTAVTSL-LPVTFTGETSAAIC-LPSVKPVVS 423
Db 1449 QGFFP-----TQTLSTPASSLVPTPAQTLSLAPGPPGLPTQTLSPAPPLAP---- 1497
QY 424 FCWDHICKPVIGTPV-----QIKLAQPGPVLVS---OPAGIPT-GSSSKQLFSLFH 469
Db 1498 -----ASPVGPAPAHITLAPASSASLLAPASVQTLTSLPAPVPTLGPAAQTALAP 1551

QY 470 V-VQPSGGNEKQVTTISHSTLTIOKCGQKTMVNTI--IPTSQPPASI 517
Db 1552 ASTQSPA-----SQASSLVWSASGAAPLPVTMWSRLPVSKDEPDTL 1592
RESULT 11
US-09-579-181-1
; Sequence 1, Application US/09579181
; Patent No. 6365372
; GENERAL INFORMATION:
; APPLICANT: Chrivia, John
; APPLICANT: Yaciuk, Peter
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247
; CURRENT APPLICATION NUMBER: US/09/579,181
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,620
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3118
; TYPE: PRT
; ORGANISM: Human
US-09-579-181-1

Query Match 4.6%; Score 197; DB 4; Length 3118;
Best Local Similarity 23.7%; Pred. No. 1.6e-06;
Matches 140; Conservative 75; Mismatches 240; Indels 136; Gaps 32;
QY 2 TLVTKVAPVSAPPKVSQG-----PRLPAPQIVAVKAPNTTTIQFPANLQL-- 46
Db 1209 TPVPLAPAPRPP--SSGLPAVLNPRPTLTGRLPTTLGTARAPMPTTLVRPLKLKLVH 1266
QY 47 --PPGTVLKNSGGLMLVSPQQTVTRAETTSNITSRPAVPANPQVVKICTVPSNSSQLI 104
Db 1267 SPSEFVSASAGNAPLTISPLHVP-----SSLPQSPASPM-----PIPNSSPLAS 1312
QY 105 KKVAVTPVKKLAQIGTVTVTVTPKPSVQSVAVPTSV-VTVTPGKP--LNTVT-TLKPS 160
Db 1313 PVSSTVSFVPLSSSLPISVPTTLAPASA-PLTIPISAPLTVSASGALLTSVTPPLAPV 1371
QY 161 LGASSTPSNEPNLKAENSAVQINL--SPTMLENVKCKNFAMLIKLCACSGSQSPMGQ 218
Db 1372 PAAPGPPLQPSGASPSALTLGLATAPSLSSQTPGHPLL-----LAPTSSHVPGLNS 1426
QY 219 NVKKLVEQ-LLDAKIEAEFEFRKLYVELKSSQPH-----LVPFLKKSVMALRQLLPS 271
Db 1427 TVAPACSPVLVPASALASPF-----PSAPNPAQAASLLAPASSASOALATPLAPMA 1478
QY 272 QSFIOQCVQQTSDMVIATCTTTVTTSVVTTTVSSSQSEKSIIVSGAT----- 320
Db 1479 AP--QTAILAPSPAPPLAP-LPVLAPSPGAAPVLASSQTPVPVMAFSPSTPGTSLASAPV 1535
QY 321 -APRTV-----SVQTL--NPLAGPVGAKAGVTVTLHVSQGTAA--TGCT-----TAGTG 365
Db 1536 PAPTVPVLAPSSQTMLPAPVPSPLPSPASTQTL-ALAPALAPTLGGSSPSQTLSLGTGNP 1594
QY 366 QTSKPLVTVANTVTVSVLSQEPKPVVSGTAVTSL-LPVTFTGETSAAIC-LPSVKPVVS 423
Db 1595 QGFFP-----TQTLSTPASSLVPTPAQTLSLAPGPPGLPTQTLSPAPPLAP---- 1643
QY 424 FCWDHICKPVIGTPV-----QIKLAQPGPVLVS---OPAGIPT-GSSSKQLFSLFH 469
Db 1644 -----ASPVGPAPAHITLAPASSASLLAPASVQTLTSLPAPVPTLGPAAQTALAP 1697
QY 470 V-VQPSGGNEKQVTTISHSTLTIOKCGQKTMVNTI--IPTSQPPASI 517
Db 1698 ASTQSPA-----SQASSLVWSASGAAPLPVTMWSRLPVSKDEPDTL 1738
RESULT 12
US-08-362-525-22

Sequence 22, Application US/08362525
Patent No. 6027910
GENERAL INFORMATION:
APPLICANT: KLAS, FRANCISCUS M.
APPLICANT: SCHREUDER, MAARTEN P.
APPLICANT: TOSCHKA, HOLSER Y.
APPLICANT: VERRIPS, CORNELIS T.
TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE
CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
PROTEIN
TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
PROTEIN
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBAY & CUSHMAN, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D. C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/362,525
FILING DATE: 04-JAN-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: EP 92202080.5
FILING DATE: 08-JUL-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: EP 92203899.7
FILING DATE: 14-DEC-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/01763
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 213289/T7020(V)
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 894 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-362-525-22

Query Match 4.5%; Score 191.5; DB 3; Length 894;
Best Local Similarity 20.1%; Pred. No. 6.1e-07;
Matches 123; Conservative 87; Mismatches 232; Indels 171; Gaps 25;

QY 22 LPAPQIVAVKAPNTTTTQFPANLQLPCTVLIKNSGPMPLVSPQ-----TVTRA 73
Db 265 VPDPNSYAVSTTTT-----EPWGTFTSTSTMTVTGNGVPTDEVIVIRTP 315
QY 74 TTSNITSRPVP-----ANPQVTKICTVPNSSQLIKKVAV---TPVKKLAQIGTVVTV 126
Db 316 TSEGLISTTTTPTGTTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 371
QY 127 PKSSVQSVAVPTSVWTVTPCKPLN-TVTTLK-PSSIGASSTPSNEP-----NLKAENS 178
Db 372 PWTGFTTSTSTMTVTGNGQPTDEVIVIRTPSTSEGLVTT-TTEPWTGFTTSTSTMS 430
QY 179 AAVQINLSPMLNKKKFLAMLIKACSGSPGNGVKKLVQLLDKAEAEFT 238
Db 431 TVTGTNGLPTD-ETVIVVKT-PTTAISSLSSSSSGGITSSI-----470
QY 239 RKLVELKSSPQHPFLVFLAKSVVALRQLPNSQSFQQCVQQTSSDVIATCTTV-TT 297

Db 471 -----TSSRPITTFYPS-----NGTSVI-----SSSVISSSVTSSTLFTS 505
QY 298 SPVVTTVTVSS-----SQSEKSIIV-----SGATAPRTVSVQTLNPLAGPVGAKA 341
Db 506 SPVSSSVISSTTST 558
QY 342 GVTVLHVGTAAGTGGTAGLLQTSKPLVTSVANTVTVSVLQPEKPVVSGTAVTSLP 401
Db 559 FISSESKSPITYS-----SSLPLVTATTSTQETAS-----SLP 592
QY 402 AVTFGETSGAAICLPKVPVSVFCWDHIC-----KPVIGTPVQIKLAQPGVLSQPAIGPT 457
Db 593 PATTKTISEQTLV-----TVSCSHVCTESISPAIVSTATVIV-----SGVTT 637
QY 458 GSSSKQLFSLPHVVQPSGGNEKQVTVTISHSTLTQKCGKTMVNTIITPSOFFPASI 517
Db 638 EYTWCPISSTETTQTKGTTEQTTETTKQTVTVTSSC-ESDVCSTKASPATVSTSTAT 696
QY 518 LKQITLPGNKILSLQASPTOKNRKENVTSCFRDEDDINDVTSMAGVNLNEENACILATN 577
Db 697 INGVTTEYTWCPISSTTESRQOTTLVTVTS-----ESGVCSETAS 737
QY 578 SELVGTLIQCKD 590
Db 738 PAIVSTATATVND 750

RESULT 13
US-08-971-692-15
Sequence 15, Application US/08971692
Patent No. 6114147
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Immobilized proteins with specific binding
TITLE OF INVENTION: capacities and their use in processes and products.
NUMBER OF SEQUENCES: 40
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/971,692
FILING DATE:
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 894 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-971-692-15

Query Match 4.5%; Score 191.5; DB 3; Length 894;
Best Local Similarity 20.1%; Pred. No. 6.1e-07;
Matches 123; Conservative 87; Mismatches 232; Indels 171; Gaps 25;

QY 22 LPAPQIVAVKAPNTTTTQFPANLQLPCTVLIKNSGPMPLVSPQ-----TVTRA 73
Db 265 VPDPNSYAVSTTTT-----EPWGTFTSTSTMTVTGNGVPTDEVIVIRTP 315
QY 74 TTSNITSRPVP-----ANPQVTKICTVPNSSQLIKKVAV---TPVKKLAQIGTVVTV 126
Db 316 TSEGLISTTTTPTGTTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 371
QY 127 PKSSVQSVAVPTSVWTVTPCKPLN-TVTTLK-PSSIGASSTPSNEP-----NLKAENS 178
Db 372 PWTGFTTSTSTMTVTGNGQPTDEVIVIRTPSTSEGLVTT-TTEPWTGFTTSTSTMS 430
QY 179 AAVQINLSPMLNKKKFLAMLIKACSGSPGNGVKKLVQLLDKAEAEFT 238
Db 431 TVTGTNGLPTD-ETVIVVKT-PTTAISSLSSSSSGGITSSI-----470

QY 239 RKLYVELKSPQPHLYPFLKSWALRQLLPNSQFIQOCVQOTSSDMVIATCTTV-TT 297
Db 471 -TSSRPITPPFPK-NGTSVI-SSSVISSSVTSSLFSS 505
QY 298 SPVVTTVSS--SOSEKSIIV--SGATAPRTVSQTLNPLAGPVGAKA 341
Db 506 SPVISSSVISSSTSTSFSSSKSVIPTSSSTSGSESETS--AGSVSSSS 558
QY 342 GVVTLHSGPTAAGTTAGTGLLOTQSKPLVTSVANTVTVLSIQPKPVVSGTAVTSLSP 401
Db 559 FISSESSKPTVSS--SSPLVTSATTSQETAS--SLP 592
QY 402 AVTGTGTSAAICLPVKPVWSECDHIC--KPVIGTPVQIKLAQPGVLSQAPAGIPT 457
Db 593 PATTTTTSQTTLV--TVTSCESHVCTESISPAIVSTATVTV--SGVTT 637
QY 458 GSSSKLFSLHVVQOPSGNGKQVNTISHSSTLTATQKCGKTMPTVNTIIPTSQPPASI 517
Db 638 EYTWCPISSTETTKTKTGTTEOTTETTKQTTVTVTSSC-ESDVCSTASPAIVSTSTAT 696
QY 518 LKQITLPGNKILSIQASPTQKNRIKENVTSCFRDEDDINDVTSMAGVNLNEENACILATN 577
Db 697 INGVTTEYTTWCPISTTESRQQTLLVTVTSC-----ESGVCSETAS 737
QY 578 SELVGTLIQSKD 590
Db 738 PAIVSTATATVND 750

RESULT 14

US-08-325-267A-2

; Sequence 2, Application US/08325267A

; Patent No. 5585271

; GENERAL INFORMATION:

; APPLICANT: WATARI, JUNJI

; APPLICANT: TAKATA, YOSHIHIRO

; APPLICANT: OGAWA, MASAHIRO

; APPLICANT: PENTTILA, MERJA

; APPLICANT: ONNELA, MAIJA-LEENA

; APPLICANT: KERANEN, SIRKKA

; TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT

; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400

; CITY: ARLINGTON

; STATE: VA

; COUNTRY: USA

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/325,267A

; FILING DATE: 18-NOV-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP PCT/JP94/00290

; FILING DATE: 24-FEB-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 38871/1993

; FILING DATE: 26-FEB-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: OBLON, NORMAN F

; REGISTRATION NUMBER: 24,618

; REFERENCE/DOCKET NUMBER: 2589-023-0XPCT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-413-3000

; TELEFAX: 703-413-2220

; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1537 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-325-267A-2

Query Match 4.3%; Score 184.5; DB 1; Length 1537;

Best Local Similarity 19.3%; Pred. No. 5.4e-06;

Matches 134; Conservative 104; Mismatches 257; Indels 201; Gaps 27;

QY 50 TVLI--KNSGFLMLVSPQQTIVTRAETTSNITSPAPVAPNP--QTVKICTVPSNSSL 104

Db 938 TVIVIRTPSEGLISTTTPWTGCTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 993

QY 105 KKVAVTPVKKLAQIGTVTVTPKSSVQSVAVPTSVTVTPGKPLN-TVTTLK-PSSLG 162

Db 994 -----GLISTTTEPWTGCTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1039

QY 163 ASSTPSNEP-----NLKAENSAAVOINLSPMLNKKKFLAMLIKACSGSOSPEM 216

Db 1040 LVTT-TTEPWTGCTST 276

QY 217 GONVKKLVEQLLDAKIEAEFTKLYVELKSSPQHPVLPFLKKSVVVALRQLLPNSQFIQ 276

Db 1097 TSSI-----TSSRPITFPYPS-----NGTSVI- 1119

QY 277 QCVOQTSSDMVIATCTTV-TTSPVTTTVSS-----SQSEKSIIV-----SGA 319

Db 1120 -----SSSVISSSVTSSLSFTSPVISSSVISSTTSTSTSTSTSTSTSTSTSTSGS 1173

QY 320 TAPRTVSQTLNPLAGPVGAKAGVTVLHSGVPTAAGTGAGTGLLQTSKPLVTSVANTV 379

Db 1174 SESETSS-----AGSVSSSFISSESSKSPYSS-----SSLPLVTSATTSQ 1215

QY 380 TTVSLOPEKPVVSGTAVTSLPAVTFGETSGAAICLPVSVKPVVFCWDHIC-----KPVIG 435

Db 1216 ETAS-----SLPPATTTKTSQTTLV-----IVTSCESHVCTESISPAIV 1255

QY 436 TPVQIKLAQPGVLSQAPIGPTGSSSKQLFSLHVVQVQPSGNGEKQVTVTSHSSTLTIOK 495

Db 1256 STATVTV-----SGVTTEYTTWCPISTTETTKQTKGTEQTTETTKQTTVTVTSS 1305

QY 496 C-----GQKTMVNTI-----IPTSOFPASILKQITLPCNKILSLQ 532

Db 1306 CESDVCSTASPAIVSTSTATINGVTTEYTTWCPISTTTSRQQTLLVTVTSCSGVCSET 1365

QY 533 ASP-----TQKNRIKENVT-----SCFRDEDDINDVTSMAGVNLNEENACILATNSELV 581

Db 1366 ASPAIVSTATATVNDVTVVPTWRPQTANEESVSSKMNSTATGETTTTLAAETTTNTVAA 1425

QY 582 GTLIQSKDEPFLFGALQKR-----ILDKKKHIDITELNSDAVN 621

Db 1426 ETITNTGAAETKTVTSSLSRGNHAETOTASATDVIGHSSSVSVSETGNTKSLTSSGLS 1485

QY 622 LISQATQERLGLLEKLTATIAQHRMTTYKASENYIL 657

Db 1486 TMSQPRSTPASSWVGYST-ASLEISTYAGSANSLL 1520

RESULT 15

US-08-700-651-5

; Sequence 5, Application US/08700651B

; Patent No. 6015882

; GENERAL INFORMATION:

; APPLICANT: PETERSEN, CAROLYN

; APPLICANT: LEECH, JAMES

; APPLICANT: NELSON, RICHARD, C.

; APPLICANT: GUT, JIRI

; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS

; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum

; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1721
; TYPE: PRT
; ORGANISM: Cryptosporidium parvum
; US-08-700-651-5

Query Match 4.28; Score 178; DB 3; Length 1721;
Best Local Similarity 20.08; Pred. No. 2.2e-05;
Matches 137; Conservative 82; Mismatches 293; Indels 172; Gaps 29;

QY	35	TTTITQFPANLQLPPTGTVLIKSNGLMLVSPQQTVTRAETTSNITSRPAVPANPQTVKIC	94
DB	286	TT	345
QY	95	TVPNSSOLIKKVAIVPKLAQIGTIVTVPKSSVQSVAVPTSVTVTPGKPLNVT	154
DB	346	TT	398
QY	155	TLKPSSLGASSTPS-----NEPNLKAENSA-----VOINLSPTMLN--	192
DB	399	TTATTTTTTSETSVIRPDEMCWLEKNGECEAKGATVGVIGDKGRIENGMAFTMIPND	458
QY	193	-----VKKCNFLAMLKILACSGSPGQNGQNVKVLVEQLLDKIE--AEFTRKLY	242
DB	459	THVFRFKVDVGTIISVRCRKGAGKLEFP-----DRSLDFTIPPVAGHNSCII	508
QY	243	VELKSSQPHLVPFLKKSVALROLLPNSQSF-----IQOCV-----QQTSSDMVIA	289
DB	509	VGVSQDGIHVSPYSGKDVSLISAPIQPSSELFNEVYCDTCTAKYGAHSGYQTSADFVTT	568
QY	290	TC-----TTVTTSPTVTTVSSSQSEKSLIVSGATAPRTVSVQTLNPLAGPVGAKGVVTL	346
DB	569	TTAKPTTTTTGAPGQPTTTTGSPSKP--TTTTTTKATTTTTTLNP-----IITT	616
QY	347	HSVGPATAGTTAGTGLLOTSKPLVTSVANTVTVSLQPEKPVV-----SGTAVTLSL	400
DB	617	TTQKPTTTT--TTKVPG-----KP--PIATTTTL-----KPIVTTTTKATTTTTTV	661
QY	401	PAVTFGETSGAAICLPSPVSVFCWDHICKPVIG-----TPVQIKLAQPGPVLSQAPGIP	456
DB	662	PTTT--TTTKRDEMTTTTTTL-----PDIGDIEITPIE-----	694
QY	457	TGSSSKOLFSLFHVQPSGGNEKQVTTISHSSTLTQKCGQKMPVNTIIPTSQFPAS	516
DB	695	-----KMLDKYTRMIYDYSNG-----LLDSNDEPIGSOAG-----IADTSNLFVQ	738
QY	517	ILKQITLPGNKLISLQASPTQKNRI-----KENVTSCFDEDDIN--D	557
DB	739	THKSTGLPIDPWGLPDPKSGNLVHPYTNOTMSGLSVSYLAARNLTV---DDETGYGLP	795
QY	558	VTSMAGVNLNEENACILATNSLYGTLIQCKDEFFLFIGALQKRILDKKKHDIETLNS	617
DB	796	IDTLTGYP--DPVSLIPFNPPE-IGELFDPISDE--IMNGTIAGIVSGISASESLLSQS	850
QY	618	DAVNLIQATQERLURGLEKLTAI	641
DB	851	ALIDPATMVMVGEFGLINPATGV	874

Search completed: February 16, 2003, 22:02:26
Job time : 40.8929 secs

This Page Blank (uspto)

GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2003, 22:29:40 ; Search time 6475 Seconds
(without alignments)
11497.294 Million cell updates/sec

Title: US-09-763-909-1
Perfect score: 2558
Sequence: 1 gggacctgtgacaaagt.....ctttattacttcttacctat 2558

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pin.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2534	99.1	2556	6	I81227	I81227 Sequence 1
2	2534	99.1	2556	9	HSTAF11	Y09321 H. sapiens T
3	1145.2	44.8	182805	9	AC017007	AC017007 Homo sapi
C 4	580	22.7	119081	2	AC121320	AC121320 Homo sapi
C 5	580	22.7	169240	9	AC022069	AC022069 Homo sapi
6	580	22.7	190782	2	AP002752	AP002752 Homo sapi
7	338.2	13.2	3603	6	I23466	I23466 Sequence 15
8	338.2	13.2	3603	6	I45759	I45759 Sequence 15
9	338.2	13.2	4233	9	HSU75308	U75308 Human Tbp-a
10	337.2	13.2	3252	9	HSTAF113	V11354 H. sapiens m
11	325.8	12.7	2196	10	AY038601	AY038601 Mus muscu
C 12	294.6	11.5	174222	2	AP001197	AP001197 Homo sapi
13	294.6	11.5	175553	2	AP001096	AP001096 Homo sapi
14	294.6	11.5	182884	9	AC007996	AC007996 Homo sapi
C 15	294.6	11.5	190782	2	AP002752	AP002752 Homo sapi
C 16	148.6	5.8	333300	2	AC125091	AC125091 Mus muscu
17	148	5.8	16903	2	AC016839	AC016839 Homo sapi
C 18	148	5.8	61682	2	AC118057	AC118057 Homo sapi
C 19	147	5.7	277892	2	AC125105	AC125105 Mus muscu
C 20	146.6	5.7	205466	2	AC127767	AC127767 Rattus no
21	143.8	5.6	3535	3	S63550	S63550 transcripti
22	142.2	5.6	3745	3	AY069807	AY069807 Drosophi
23	142.2	5.6	4615	3	DROTAFL110X	L06861 Drosophila
24	142.2	5.6	4615	6	I23459	I23459 Sequence 1
25	142.2	5.6	4615	6	I45752	I45752 Sequence 1
26	137.6	5.4	205466	2	AC127767	AC127767 Rattus no
27	124.4	4.9	62974	2	AC118056	AC118056 Homo sapi
28	121.8	4.8	277892	2	AC125105	AC125105 Mus muscu
29	112.4	4.4	158509	2	AC114177	AC114177 Rattus no
C 30	109.8	4.3	62974	2	AC118056	AC118056 Homo sapi
C 31	99	3.9	158509	2	AC114177	AC114177 Rattus no
C 32	98.4	3.8	191918	2	AC102692	AC102692 Mus muscu
C 33	94.6	3.7	121982	9	HS1107C24	AL109911 Human DNA
C 34	92	3.6	333300	2	AC125091	AC125091 Mus muscu
C 35	91.2	3.6	64035	2	AC119227	AC119227 Mus muscu
C 36	89.6	3.5	194426	2	AL663067	AL663067 Mus muscu
C 37	71	2.8	112277	2	AC113239	AC113239 Tetraodon
C 38	62.4	2.4	2147	9	AK097744	AK097744 Homo sapi
C 39	58.6	2.3	59865	2	AC014422	AC014422 Drosophi
C 40	58.6	2.3	168479	3	AC093454	AC093454 Drosophi
C 41	58.6	2.3	176056	3	AC010066	AC010066 Drosophi
C 42	58.6	2.3	283821	3	AE003528	AE003528 Drosophi
C 43	57.6	2.3	7218	6	I66494	I66494 Sequence 14
C 44	56.8	2.2	126712	9	AL137077	AL137077 Human DNA
C 45	56.4	2.2	201470	2	AC113649	AC113649 Rattus no

ALIGNMENTS

RESULT 1	I81227	I81227	2556 bp	DNA	linear	PAT 10-JUN-1998
LOCUS	I81227	Sequence 1 from patent US 5710025.				
DEFINITION	I81227					
ACCESSION	I81227					
VERSION	I81227.1	GI:3209517				
KEYWORDS						
SOURCE		Unknown.				
ORGANISM		Unclassified.				
REFERENCE		1 (bases 1 to 2556)				
AUTHORS		Dikstein,R. and Tjian,R.				
TITLE		Cell-type specific transcription factor				
JOURNAL		Patent: US 5710025-A 1 20-JAN-1998;				
FEATURES		Location/Qualifiers				

[illegible]

QY	961	AGCACCCAGAACTGTGTCACTGTCAGTGCAGAACTTTGAACCCCACTTGTGTGTCAGTGGGAGCAAA	1020
DB	960	AGCACCCAGAACTGTGTCACTGTCAGTGCAGAACTTTGAACCCCACTTGTGTGTCAGTGGGAGCAAA	1019
QY	1021	AGCTGGAGTTGTGACACTTTCATTTCTGTGGGCCCACTGCTGCAACAGAGGAGCAACAACG	1080
DB	1020	AGCTGGAGTTGTGACACTTTCATTTCTGTGGGCCCACTGCTGCAACAGAGGAGCAACAACG	1079
QY	1081	TGGAACTGGTTTGGCTTCAGACTTTCAAAACCACTTGTGACATCTGTGGCAACACACAGTGCAC	1140
DB	1080	TGGAACTGGTTTGGCTTCAGACTTTCAAAACCACTTGTGACATCTGTGGCAACACACAGTGCAC	1139
QY	1141	CACGGTCTCACATGCCAACCTGAAAGCCAGTTGTCTCTGGAACAGCAGTAACACTGTCCCT	1200
DB	1140	CACGGTCTCACATGCCAACCTGAAAGCCAGTTGTCTCTGGAACAGCAGTAACACTGTCCCT	1199
QY	1201	TCCAGCAGTAACATTTTGGAGAACTTCAGGTGCGAGTATTTGCTTCCATCTGTGAAACC	1260
DB	1200	TCCAGCAGTAACATTTTGGAGAACTTCAGGTGCGAGTATTTGCTTCCATCTGTGAAACC	1259
QY	1261	TGTTGTTTCCCTCTGCTGGGACCACATCTGCAAGCCTGTTATTGGAGCTCCAGTTCAAAT	1320
DB	1260	TGTTGTTTCCCTCTGCTGGGACCACATCTGCAAGCCTGTTATTGGAGCTCCAGTTCAAAT	1319
QY	1321	CAAACTTGGCCAGCGGGCCCTGTCTTTTCAACAACAGCTGGGATTTCCAACAGGCAGTTC	1380
DB	1320	CAAACTTGGCCAGCGGGCCCTGTCTTTTCAACAACAGCTGGGATTTCCAACAGGCAGTTC	1379
QY	1381	AAGCAAGCACTATTCTCATTTGTTTCAAGTGTGACAGCCCTTCAGGAGGCATGAAAA	1440
DB	1380	AAGCAAGCACTATTCTCATTTGTTTCAAGTGTGACAGCCCTTCAGGAGGCATGAAAA	1439
QY	1441	ACAAGTGACCACAATTTTCACATTCCTCAACATTGACCATTCAGAAATCTGGACAGAGAC	1500
DB	1440	ACAAGTGACCACAATTTTCACATTCCTCAACATTGACCATTCAGAAATCTGGACAGAGAC	1499
QY	1501	GATCCAGTGAACACCACTAATACCTACTAGTCAGTTTCCCTCCAGCTTCCATCTTAAAGCA	1560
DB	1500	GATCCAGTGAACACCACTAATACCTACTAGTCAGTTTCCCTCCAGCTTCCATCTTAAAGCA	1559
QY	1561	AATTTACCTCTGCCTGGAAATAAATTTCTGTCACTTCAAGCATCTCCTACTCAGAAAAATA	1620
DB	1560	AATTTACCTCTGCCTGGAAATAAATTTCTGTCACTTCAAGCATCTCCTACTCAGAAAAATA	1618
QY	1621	GAATTAAGAGATGTAAACATCATGCTCCGAGATGAGGATGACATCAATGATGTGACTT	1680
DB	1619	GAATTAAGAGATGTAAACATCATGCTCCGAGATGAGGATGACATCAATGATGTGACTT	1678
QY	1681	CTATGGCAGGGGTCAACCTTTAATGAAGAAATGCCCTGCACTTTAGCAACAACTCTGAAT	1740
DB	1679	CTATGGCAGGGGTCAACCTTTAATGAAGAAATGCCCTGCACTTTAGCAACAACTCTGAAT	1738
QY	1741	TGGTTGGCAGCACTTCATGCTCATGTAAAGATGAACCATTTCTTTTATTGGAGCTCTAC	1800
DB	1739	TGGTTGGCAGCACTTCATGCTCATGTAAAGATGAACCATTTCTTTTATTGGAGCTCTAC	1798
QY	1801	AAAAAGAACTTTAGACATTTGGTAAAAAGCATGACATACAGAACTTAACTCTCATGCTG	1860
DB	1799	AAAAAGAACTTTAGACATTTGGTAAAAAGCATGACATACAGAACTTAACTCTCATGCTG	1858
QY	1861	TGAACCTTGATCTCCCAAGCAACACAGAAACGACTACGAGGCCCTTCAGAAAACTGACTG	1920
DB	1859	TGAACCTTGATCTCCCAAGCAACACAGAAACGACTACGAGGCCCTTCAGAAAACTGACTG	1918
QY	1921	CAATTTGCTCAGCATCGAATGACTTACAGGCAAGTGAAATTTACATCTCTGTGTAGTG	1980
DB	1919	CAATTTGCTCAGCATCGAATGACTTACAGGCAAGTGAAATTTACATCTCTGTGTAGTG	1978
QY	1981	ATACCAGTCCAGCTCAAAATTTCTTGAAGAGCTGGATCAATTTGGAGAGCAGAGAAAGG	2040
DB	1979	ATACCAGTCCAGCTCAAAATTTCTTGAAGAGCTGGATCAATTTGGAGAGCAGAGAAAGG	2038
QY	2041	ATTTTGAAGAAAGAAATGTTTACTTTTAAAGGCACCCAAAGAGTCGTTCTTAATAAAGAAATC	2100

```
|||||
Db 2039 ATTTGGAGAAAGAGAAATGTTACTTAAGCAGCCAGAGTCGTTCTTAATAAGAAAGATC 2098
QY 2101 CAGAACAGCTGAGATTAAAGCAGAAAGCCAAAGAGTTACAGCAATTTGGAACCTTGACACAGA 2160
Db 2099 CAGAACAGCTGAGATTAAAGCAGAAAGCCAAAGAGTTACAGCAATTTGGAACCTTGACACAGA 2158
QY 2161 TACAGCATAGAGAGCTTAATCTACAGCTCTGCGAGCTATTGGACCAAGGAAGAGAGAC 2220
Db 2159 TACAGCATAGAGAGCTTAATCTACAGCTCTGCGAGCTATTGGACCAAGGAAGAGAGAC 2218
QY 2221 CACTAGAACTTGAATTTGAGGGCTTTAAAGACAAACCTCTTCTGCTTCTGGACATCCAGCC 2280
Db 2219 CACTAGAACTTGAATTTGAGGGCTTTAAAGACAAACCTCTTCTGCTTCTGGACATCCAGCC 2278
QY 2281 TGACAGCCCAACACAGTTGATGCTGCTCCAAAGAAATCAGAGAAATCTGCGCTCAGGGACTTGA 2340
Db 2279 TGACAGCCCAACACAGTTGATGCTGCTCCAAAGAAATCAGAGAAATCTGCGCTCAGGGACTTGA 2338
QY 2341 TATTTTCTATGACAGAGAGAGGAGATGAATATTCAGAGCTCTATACCTGGCCCTTC 2400
Db 2339 TATTTTCTATGAGAACAGAGAGGAGATGAATATTCAGAGCTCTATACCTGGCCCTTC 2398
QY 2401 TGAAGTGACCACTCCACTCTTCCATCCACATCTCTTGCTATTACTGCGCAAGAGAGACACA 2460
Db 2399 TGAAGTGACCACTCCACTCTTCCATCCACATCTCTTGCTATTACTGCGCAAGAGAGACACA 2458
QY 2461 AAGCATTGTTGACATGCTGCTGAAATTTCAATTTCTTGGAATAAACCAACATGAAGAG 2520
Db 2459 AAGCATTGTTGACATGCTGCTGAAATTTCAATTTCTTGGAATAAACCAACATGAAGAG 2518
QY 2521 CATGTTTACGATTAGACATTTTAACTCTTACCTAT 2558
Db 2519 CATGTTTACGATTAGACATTTTAACTCTTACCTAT 2556

RESULT 2
HSTAFII
LOCUS HSTAFII 2556 bp mRNA linear PRI 14-NOV-1996
DEFINITION H.sapiens TAFII105 mRNA, partial.
ACCESSION Y09321
VERSION Y09321.1 GI:1669688
KEYWORDS TAFII105 gene.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Dikstein,R., Zhou,S. and Tjian,R.
TITLE Human TAFII 105 is a cell type-specific TFIID subunit related to
hTAFII130
JOURNAL Cell 87 (1), 137-146 (1996)
MEDLINE 97011146
PUBMED 8858156
REFERENCE
AUTHORS Dikstein,R.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-1996) R. Dikstein, Weizmann Institute of Science,
Dept. Biochemistry, Rehovot, 76100, ISRAEL
FEATURES
source
1. .2556
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HeLa"
1. .2406
/gene="TAFII105"
<1. .2406
/gene="TAFII105"
/codon_start=1
/product="TBP associated factor"
/protein_id="CAA70499.1"
/db_xref="GI:1669689"
/db_xref="SWISS-PROT:Q92750"
```

```
/translation="GTLVTKVAPVSAPPKVVSSGPRLPAPQIVAVKAPNTTITQFPANL
QLDPGVLIKNSGPMMLVSPQQTIVTARETTISNTRSPAVPAMPQIVKICTVTPNSSSQ
LIKRVAVTPVKLAQIGTTTVPKPSVQSVAVTVPVTPGPKLNTVTKPSS
LGASSTPSNEPLNKAENSAQINLSPMLNKKCNFLKMLIKLACSGSQSPMGQ
NVKLLVQQLDLDAKEAEFTKLYVELKSPQPHLVFLKSVVALRQLLPSQSFIO
QCVQOTSSDMVIATCTTVTTSPTVTSSVSSOSEKSIIVSGATPRTVSVOTPLNLA
GPKAKAGVVTLSVGPRTAGTGTAGTLLQTSKPLVTSVANTVTVSLQPKPVVS
GTAUTLSLPAVTEGSGAAICULPSVPVSVFCDWHICKPVGTPVOIKLQAPVLS
QPAIGTSSSKQLFSLFHVQPSGNEKQVTTISHSSTLTQKCGQKMPYNTIIP
TSQFPASILKQITLPGNKILSLQASPTQKNRIKENVTSFREDIDNDVTSAGVNL
NEENACILATNSELVGLTQSCDEPFLFYGALQKRLIDIKKKHDTIELNSDVKDLIS
OATQEBRLGLLEKLTALAOHRMTTYKASENYILCSDTRSOLKFLKLELDKQKDLK
ERMLLKAASRSNKEDEPQLRLKOKAKELQLELAQIQRDANLALAAIGPRKRP
LSSGIGLKNLLASGTSLSLTATQQLHRPRITRILCDLILFCMEQREMKYSLALYA
LK"
BASE COUNT 797 a 614 c 514 g 631 t
ORIGIN
Query Match 99.1%; Score 2534; DB 9; Length 2556;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2556; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
QY 1 GGGACCCCTGGTGACAAAGTGGCTCCGGTCAGCGCCCTCCTAAAGTCAGCAGCGCCCT 60
Db 1 GGGACCCCTGGTGACAAAGTGGCTCCGGTCAGCGCCCTCCTAAAGTCAGCAGCGCCCT 60
QY 61 AGGCTGCTCTCTCTCAGATAGTGGCGCTGAAAGCCCCCAACACACAGCAATCCAGTTT 120
Db 61 AGGCTGCTCTCTCTCAGATAGTGGCGCTGAAAGCCCCCAACACAGCAATCCAGTTT 120
QY 121 CCTGCTAATTTGACGCTCTCTCAGGAACCGTTTGTATTAAGTAACAGTGGTCCGTTG 180
Db 121 CCTGCTAATTTGACGCTCTCTCAGGAACCGTTTGTATTAAGTAACAGTGGTCCGTTG 180
QY 181 ATGTTGGTATCTCTCTCAGCAAACTGTAAAGAGCCGAGACCAAGTAACATAACCTCA 240
Db 181 ATGTTGGTATCTCTCTCAGCAAACTGTAAAGAGCCGAGACCAAGTAACATAACCTCA 240
QY 241 AGGCCAGCAGTACAGCGAATCTCTCAAAAGTCAAAATCTGTACAGTGGCGAATCTAGC 300
Db 241 AGGCCAGCAGTACAGCGAATCTCTCAAAAGTCAAAATCTGTACAGTGGCGAATCTAGC 300
QY 301 TCACAATTAATCAGAAAGTGGCAGTGCACACCTGTTTAAAAAATTTGGCAAAATAGGA 360
Db 301 TCACAATTAATCAGAAAGTGGCAGTGCACACCTGTTTAAAAAATTTGGCAAAATAGGA 360
QY 361 ACTGTGTAAACCACTGTTTCCGAAGCCCTCTCTCAGTACAAATCTGTGGCTGTGCCAAC 420
Db 361 ACTGTGTAAACCACTGTTTCCGAAGCCCTCTCTCAGTACAAATCTGTGGCTGTGCCAAC 420
QY 421 RGTGTCACAGTACTCTCTGGAAGCCATTGGAATPACTGTAACTACCTGAGCGCTTCAAG 480
Db 421 -GTGTCACAGTACTCTCTGGAAGCCATTGGAATPACTGTAACTACCTGAGCGCTTCAAG 479
QY 481 TTTGGGAGCATCTCCACTCTTCAAAATGAGCCCAATCTTAAAGCAGAGAACTCAGCAGC 540
Db 480 TTTGGGAGCATCTCCACTCTTCAAAATGAGCCCAATCTTAAAGCAGAGAACTCAGCAGC 539
QY 541 TGTTCAGATTAATCTTTCTCCGACAATGTGAGAAATGTGAGAAATGTGAGAAATGTGAG 600
Db 540 TGTTCAGATTAATCTTTCTCCGACAATGTGAGAAATGTGAGAAATGTGAGAAATGTGAG 599
QY 601 TGCATATTTAATAAATAGCATGTAGTGGATCAGTCCCTCCCTGAAATGGGCAAAATGT 660
Db 600 TGCATATTTAATAAATAGCATGTAGTGGATCAGTCCCTCCCTGAAATGGGCAAAATGT 659
QY 661 GAAGAACTGTTGCAACAACCTTTTGGATGCAAAATGCAAGCAAGAAATTTACTAGAA 720
Db 660 GAAGAACTGTTGCAACAACCTTTTGGATGCAAAATGCAAGCAAGAAATTTACTAGAA 719
QY 721 ACTGTATGTTGAACCTCAAGCTTCTCACCTCAGCCCTCAGCTGGTCTCTTTCTTAAGAAA 780
Db 720 ACTGTATGTTGAACCTCAAGCTTCTCACCTCAGCCCTCAGCTGGTCTCTTTCTTAAGAAA 779
```


REFERENCE
AUTHORS
TITLE
JOURNAL

University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 182805)
Waterston,R.H.
Direct Submission
Submitted (29-MAY-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 182805)
Waterston,R.H.
Direct Submission
Submitted (08-AUG-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 182805)
Waterston,R.
Direct Submission
Submitted (10-AUG-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Aug 8, 2002 this sequence version replaced gi:21240754.

COMMENT

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics

Center project name: H_NH0119H12

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Moon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong
and coworkers at <http://www.chori.org>
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-67F3; the clone sequenced
to the right is RP11-224E18. Actual start of this clone is at base
position 1 of RP11-119H12; actual end is at base position 182805 of
RP11-224E18.

Data from AC108042 was used to finish this clone, AC017007.

FEATURES
source

Location/Qualifiers
1. 182805
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="UL"
/map="4"
/clone="RP11-119H12"
/clone_lib="RPCI-11"

repeat_region
1. 57
/rpt_family="Alu"
58. 344
/rpt_family="Alu"
345. 501
/rpt_family="Alu"
502. 627
/rpt_family="L1"
638. 942
/rpt_family="Alu"
998. 1150
/rpt_family="Alu"
1193. 1350
/rpt_family="Achoho"
1353. 1440
/rpt_family="MIR"
1443. 1734
/rpt_family="Alu"
1996. 2126
/rpt_family="Alu"
2133. 2441
/rpt_family="Alu"
2847. 2885
/rpt_family="(CAATT)n"
2966. 3270
/rpt_family="Alu"
3312. 3352
/rpt_family="(TA)n"
3877. 4024
/rpt_family="L1"
4299. 4557
/rpt_family="Alu"
4873. 5162
/rpt_family="L1"
5339. 5367
/rpt_family="AT-rich"
5377. 5677
/rpt_family="Alu"
5679. 5759
/rpt_family="MIR"
6137. 6176
/rpt_family="(TGG)n"
6206. 6427
/rpt_family="MIR"
6624. 7163
/rpt_family="MaLR"
7234. 7323
/rpt_family="L2"
7324. 7629
/rpt_family="Alu"
7630. 7663
/rpt_family="L2"
8060. 8100
/rpt_family="(CAAA)n"
8527. 8820
/rpt_family="Alu"
9005. 9210
/rpt_family="Alu"
9426. 9618
/rpt_family="Alu"
9619. 9917
/rpt_family="Alu"
9918. 9928
/rpt_family="Alu"
9929. 10191
/rpt_family="Alu"
10192. 10324
/rpt_family="Alu"
10903. 11011
/rpt_family="Alu"
11015. 11043
/rpt_family="(CAAA)n"
11206. 11495
repeat_region

/rpt_family="Alu"
11593..11619
/rpt_family="TG)n"
11923..12151
/rpt_family="Alu"
12670..13040
/rpt_family="Alu"
13065..13315
/rpt_family="Alu"
13319..13372
/rpt_family="Mariner"
13468..13545
/rpt_family="L1"
13546..13845
/rpt_family="Alu"
13846..14212
/rpt_family="L1"
14491..14727
/rpt_family="L1"
14728..15028
/rpt_family="Alu"
15029..15174
/rpt_family="L1"
15586..15732
/rpt_family="Alu"
18088..18110
/rpt_family="TTTA)n"
18114..18395
/rpt_family="Alu"
18941..19238
/rpt_family="Alu"
19266..19346

Query Match 44.8%; Score 1145.2; DB 9; Length 182805;
Best Local Similarity 91.5%; Pred. No. 1.3e-289;
Matches 1319; Conservative 0; Mismatches 98; Indels 25; Gaps 9;
QY 1086 CTGGTTTGGCTTCAGACTTC-AAAACACTGTGTGACATCTGTGGCAACACAGTGACCAG 1144
DB 95498 CTGTTTGGCTTCAGACTTCGAAACCACTGTGTGACATCTGTACCAACACAGTGACAAATG 9557
QY 1145 GTCCTACTCAACCTGAAAGCCAGTGTCTCTGACAGCAGTAACTGTCCTCCATCA 1204
DB 95558 GTCTACTCAACCTGAAAGCCAGTGTCTCTGACAGCAGTAACTGTCCTCCATCA 95617
QY 1205 GCAGTAACCTTTGGAGAACTTCAGGTGCAGCTATTGTCTCCATCTGTGAAACCTGTT 1264
DB 95618 GCAGTAACCTTTGAGAACTTCAGGTGCAGCTATTGTCTCCATCTGTGAAACCTTTT 95677
QY 1265 GTTTCCTCTGCTGGACACATCTG-CAAGCCTGTTATTTGGGACTCCAGTTCAAATCAA 1323
DB 95678 G-TTACATCTACTGGGACCACATCTGACAAGCCTGTTATTGGCACTCCCAATGCAATCAA 95736
QY 1324 ACTTCCCGACCGCGCTGCTCTTCAACACAGCTGGGATTCACAGGCGAGTTCAG 1383
DB 95737 ACTTGGCCAGCGCGCTGCTCTTCAACACAGCTGGGATTCACAGGCGAGTTCAGT 95796
QY 1384 CAAGCAACTATTCTCATTGTTTCAGCTAGTTTCAGCAGCCTTCAGGAGGCAATGAAACAA 1443
DB 95797 CAAGCAACTA-----GTAGTTCAGCAGGCTTCAGGAGGCAATGAAACAA 95841
QY 1444 AGTGACCACAATTTACATCTCAACATTTGACCAATTCAGAAATGTGGACAGACGAT 1503
DB 95842 GGTGACCACTATACATCTCCTGACATTTGACCAATTCAGAAATCTGGATAGAACAT 95901
QY 1504 GCCAGTGAACACCAATTAACCTACTAGTCACTTTCTCCAGCTTCCTCACTCAACCAAT 1563
DB 95902 GCCAGTGAACACCAATTAACCTACTAGTCACTTTCTCCAGCTTCCTCACTCAACCAAT 95961
QY 1564 TACCTCTGCTGCAATTAATTTCTGCTCACTTCAAGCATCTCTACTCAGAAATAAGAA 1623
DB 95962 TA-CTCTGCTGCAATTAATTTCTGCTCTCAAGCATCTCTACTCAGAAATAAGAA 96020

QY 1624 TAAAGAGAATGTACATCATGCTTCCGAGATGAGGATGACATCAATGATGACTTCTA 1683
DB 96021 TAAAGAGAATGGAACAATATGCTTCAGGATGAAGATGACATCAATGATGACTTCTA 96080
QY 1684 TGGCAGGGGTCAACCTTAATGAAGAAAATGCCTGCATCTTAGCAACAACCTCGAATTGG 1743
DB 96081 TGGCAGGGGTCAACCTTAATGAAGAAAATGCCTGTCTTAGCAACAACCTCGAATTGG 96140
QY 1744 TTGGCACACTCATTCAGTTCATGTAAAGATGAACCAATTTCTTTTATTGGAGCTCTACAAA 1803
DB 96141 TTGGCACACTCATTCAGTTCATGTAAAGATGAACCAATTTCTTTTATTGGAGTCTACAAA 96200
QY 1804 AGAGAATCTTAGACATTTGGTAAAAGCATGACATTACAGAATCTTAATCTGATGCTGTA 1863
DB 96201 AGAGAATTTTAGACATTTGGTAAAAGCATGACATTACAGAATCTTAATCTGATGCTGTA 96258
QY 1864 ACTTGATCTCCCAAGCAACACAGGAAAGACTAGAGGCTTCTAGAAAACCTGACGCA 1923
DB 96259 ACTTGATCAACCATGCAACACAGGAAAATTTATGAAGCCTTCTAGAAAACCTGACGCA 96318
QY 1924 TTGCTCAGCATGCAATGACTACTTACAAGCAAGTGAATTTACATCTCTGCTAGTGATA 1983
DB 96319 TTGCTCAGCATGCAATGACTACTTACAAGCAAGTGAATTTACATCTCTGCTAGTGATA 96378
QY 1984 CCAGTCAACAGCTCAAAATTTCTTGAAGCTGGATCAATTTGGAGAGCAGAGAAAGGATT 2043
DB 96379 CCAGTCAACAGCTCAAAATTTCTTGAAGCTGGATCAATTTGGAGAGCAGAGAAAGGATT 96438
QY 2044 TCGAAGAAAGAGAAATGTTACTTTAAGGCAAGCAAGTCTTCTTAATAAAGAAAGTCCAG 2103
DB 96439 TAGAAGAAAGAGAAATGTTACTTTAAGGCAAGCAAGTCTTCTTAATAAAGAAAGTCCAG 96498
QY 2104 AACAGCTGAGATTAAAGCAGAAAGCCAAAGAGTTACAGCAATTTGGAACCTTGACACATAC 2163
DB 96499 AACAGCTGAGATTAAAGCAGAAAGCCAAAGAGTTACAGCAATTTGGAACCTTGACACATAC 96558
QY 2164 AGCATAGAGACGCTAATCTCAAGCTCTTGCAGCTTATGGACCAAGGAAAGAGAGACCAC 2223
DB 96559 AGCATAGAGATGCTAATCTCAAGCTCTTGCAGCTTATGGACCAAGGAAAGAGAGACCAC 96618
QY 2224 TAGAATCTGGAATGAGGCTTTAAAGAGCAACCTTCTGCTTCTGGGACATCCAGCTGA 2283
DB 96619 TAGAATCTGGAATGAGGCTTTAAAGAGCAACCTTCTGCTTCTGGGACATCCAGCTGA 96678
QY 2284 CAGCACCACAAACAGTTGCATCGTCCAAAGATCACAGAAATCTGCCCTCAGGACTTCGATAT 2343
DB 96679 CAGTCAACCAACAGTTGCTTCATCTAAGAAATCAAGAATCTGCCCTCAGGACTTCGATAT 96738
QY 2344 TTTGATGGAACAGGAACGGAGATGAAGTATCTCGAGCTCTATACCTGGCCCTTCTGA 2403
DB 96739 TTTGATGGAACAGGAAGAGAGATGAAGTATCTCGAGCTCTATACCTGGCCCTTCTGA 96798
QY 2404 AGTGACCACTCCACTTCCATCCACATCTTGTCTATTTACTGCCCAAGCAACACAA-- 2461
DB 96799 AGTGACCACTCTACTC-TCATCCAGATCTTGCATTTACTGCCCAAGCAACACAAAG 96857
QY 2462 AGCATTTGTCAGTCTGCTGAAATTTCAATTTCTGGAATAA-CACCAACATGAAGAG 2520
DB 96858 AGCATTTGTCAGTCTGCTGAAATTTCAATTTCTGGAATAA-CACCAACATGAAGAG 96917
QY 2521 CA 2522
DB 96918 CA 96919

RESULT 4
AC121320/c
LOCUS
DEFINITION Homo sapiens chromosome 18 clone RP11-25D3 map 18, *** SEQUENCING IN PROGRESS ***, 4 unordered pieces.
AC121320
ACCESSION AC121320.2 GI:22296957
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.

LOCUS AC022069 169240 bp DNA linear PRI 16-JUN-2002
 DEFINITION Homo sapiens, clone RP11-107K17, complete sequence.
 ACCESSION AC022069
 VERSION AC022069.6 GI:21429684
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 169240)
 TITLE Homo sapiens, clone RP11-107K17
 JOURNAL Unpublished
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 REFERENCE 2 (bases 1 to 169240)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Bida, F.,
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,
 Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P.,
 DeAtellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,
 Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
 Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,
 Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
 McPheeters, R., Meldrim, J., Meneus, L., Morrow, J., Naylor, J.,
 Norman, C.H., O'Connor, T., O'Donnell, P., Oliver, T.M., Peterson, K.,
 Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,
 Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
 Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (25-JAN-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 169240)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A.,
 Cook, A., Cooke, P., DeAtellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Fato, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K.,
 Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N.,
 Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J.,
 Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
 Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
 Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
 Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,
 Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zalnoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (30-MAY-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 4 (bases 1 to 169240)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A.,
 Cook, A., Cooke, P., DeAtellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Fato, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K.,
 Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N.,
 Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J.,
 Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
 Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
 Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
 Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,
 Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zalnoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (30-MAY-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 4 (bases 1 to 169240)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A.,
 Cook, A., Cooke, P., DeAtellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Fato, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K.,
 Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N.,
 Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J.,

Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
 Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
 Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
 Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,
 Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zalnoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (01-JUN-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 5 (bases 1 to 169240)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A.,
 Cook, A., Cooke, P., DeAtellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Fato, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K.,
 Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N.,
 Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J.,
 Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
 Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
 Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
 Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,
 Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zalnoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (16-JUN-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 16, 2002 this sequence version replaced gi:21263339.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L969
 Center clone name: 107_K_17

 Location/Qualifiers
 1..169240
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-107K17"
 /clone_lib="RPCI-11 Human Male BAC"
 complement(2..164)
 /rpt_family="L3"
 165..203
 /rpt_family="(WTCA)n"
 complement(324..389)
 /rpt_family="L3"
 940..1236
 /rpt_family="AluSx"
 1657..1877
 /rpt_family="MIR"
 complement(2045..2391)
 /rpt_family="AluSx"
 complement(3134..3360)
 /rpt_family="L2"
 complement(3529..4480)

TITLE
JOURNAL

COMMENT

FEATURES
source

repeat_region
 repeat_region
 repeat_region
 repeat_region
 repeat_region
 repeat_region
 repeat_region

Quality coverage: 4.15x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 38 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```
1 14910 contig of 14910 bp in length
15011 26140 contig of 11130 bp in length
36929 contig of 10689 bp in length
37030 48911 contig of 11882 bp in length
57744 contig of 8733 bp in length
68839 contig of 10995 bp in length
77568 contig of 8629 bp in length
87657 contig of 9989 bp in length
93985 contig of 6228 bp in length
100964 contig of 6879 bp in length
107297 contig of 6233 bp in length
113948 contig of 6551 bp in length
120683 contig of 6635 bp in length
126925 contig of 6142 bp in length
130951 contig of 3926 bp in length
134620 contig of 3569 bp in length
138171 contig of 3451 bp in length
143486 contig of 5215 bp in length
148147 contig of 4950 bp in length
152197 contig of 3950 bp in length
155994 contig of 3697 bp in length
159446 contig of 3352 bp in length
162952 contig of 3406 bp in length
165556 contig of 2504 bp in length
168139 contig of 2483 bp in length
170861 contig of 2622 bp in length
173710 contig of 2749 bp in length
175823 contig of 2013 bp in length
177259 contig of 1336 bp in length
178684 contig of 1892 bp in length
180676 contig of 1490 bp in length
182266 contig of 2206 bp in length
184572 contig of 1599 bp in length
186271 contig of 1717 bp in length
188088 contig of 1125 bp in length
189189 contig of 1172 bp in length
190585 contig of 1172 bp in length
190686 190782 contig of 97 bp in length.
```

* NOTE: This is a 'working draft' sequence. It currently consists of 38 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
1 14910: contig of 14910 bp in length
14911 15010: gap of 100 bp
15011 26140: contig of 11130 bp in length
26141 26240: gap of 100 bp
26241 36929: contig of 10689 bp in length
36930 37029: gap of 100 bp
37030 48911: contig of 11882 bp in length
48912 49011: gap of 100 bp
49012 57744: contig of 8733 bp in length
57745 57844: gap of 100 bp
57845 68839: contig of 10995 bp in length
68840 68939: gap of 100 bp
68940 77568: contig of 8629 bp in length
77569 77668: gap of 100 bp
77669 87657: contig of 9989 bp in length
87658 87757: gap of 100 bp
87758 93985: contig of 6228 bp in length
93986 94085: gap of 100 bp
```

```
* 94086 100964: contig of 6879 bp in length
100965 101064: gap of 100 bp
101065 107297: contig of 6233 bp in length
107298 107397: gap of 100 bp
107398 113948: contig of 6551 bp in length
113949 114048: gap of 100 bp
114049 120683: contig of 6635 bp in length
120684 120783: gap of 100 bp
120784 126925: contig of 6142 bp in length
126926 127025: gap of 100 bp
127026 130951: contig of 3926 bp in length
130952 131051: gap of 100 bp
131052 134620: contig of 3569 bp in length
134621 134720: gap of 100 bp
134721 138171: contig of 3451 bp in length
138172 138271: gap of 100 bp
138272 143486: contig of 5215 bp in length
143487 143586: gap of 100 bp
143587 148147: contig of 4950 bp in length
148148 148247: gap of 100 bp
148248 152197: contig of 3950 bp in length
152198 152297: gap of 100 bp
152298 155994: contig of 3697 bp in length
155995 156094: gap of 100 bp
156095 159446: contig of 3352 bp in length
159447 159546: gap of 100 bp
159547 162952: contig of 3406 bp in length
162953 163052: gap of 100 bp
163053 165556: contig of 2504 bp in length
165557 165656: gap of 100 bp
165657 168139: contig of 2483 bp in length
168140 168239: gap of 100 bp
168240 170861: contig of 2622 bp in length
170862 170961: gap of 100 bp
170962 173710: contig of 2749 bp in length
173711 173810: gap of 100 bp
173811 175823: contig of 2013 bp in length
175824 175923: gap of 100 bp
175924 177259: contig of 1336 bp in length
177260 177359: gap of 100 bp
177360 178684: contig of 1325 bp in length
178685 178784: gap of 100 bp
178785 180676: contig of 1892 bp in length
180677 180776: gap of 100 bp
180777 182266: contig of 1490 bp in length
182267 182366: gap of 100 bp
182367 184572: contig of 2206 bp in length
184573 184672: gap of 100 bp
184673 186271: contig of 1599 bp in length
186272 186371: gap of 100 bp
186372 188088: contig of 1717 bp in length
188089 188188: gap of 100 bp
188189 189313: contig of 1125 bp in length
189314 189413: gap of 100 bp
189414 190585: contig of 1172 bp in length
190586 190685: gap of 100 bp
190686 190782: contig of 97 bp in length.
```

FEATURES

source

```
1..190782
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q11.2"
/clone="Rp11-802C10"
1..14910
/note="assembly_fragment"
15011..26140
/note="assembly_fragment"
26241..36929
/note="assembly_fragment"
37030..48911
/note="assembly_fragment"
49012..57744
```

misc_feature

```
/note="assembly_fragment"
```

misc_feature

```
/note="assembly_fragment"
```

misc_feature

```
/note="assembly_fragment"
```

misc_feature

```
/note="assembly_fragment"
```

misc_feature

```
/note="assembly_fragment"
```

```
misc_feature /note="assembly_fragment"
57845..68839
misc_feature /note="assembly_fragment"
68940..77568
misc_feature /note="assembly_fragment"
77669..87657
misc_feature /note="assembly_fragment"
87758..93985
misc_feature /note="assembly_fragment clone_end:77 vector_side:left"
94086..100964
misc_feature /note="assembly_fragment"
101065..107297
misc_feature /note="assembly_fragment"
107398..113948

Query Match 22.7%; Score 580; DB 2; Length 190782;
Best Local Similarity 97.3%; Pred. No. 3.5e-141;
Matches 611; Conservative 0; Mismatches 15; Indels 2; Gaps 2;

QY 769 TCCTAAGAAAGCGTGGTGGCTTACGACAACTTCTGCCTAACTCCAGAGCTTCATCCA 828
Db 166175 TTTCCAGAAAGCGTGGTGGCTTACGACAACTTCTGCCTAACTCCAGAGCTTCATCCA 166234

QY 829 GCAATGTGTTTCAGCAGACTTCTAGTGACATGGTCACTTGTACCTGTACTACACAGTAAC 888
Db 166235 GCAATGTGTTTCAGCAGACTTCTAGTGACATGGTCACTTGTACCTGTACTACACAGTAAC 166294

QY 889 AACTTCTCCTGTGGTGACAACTACAGTGTCTCTCAAGCAGTCTGTGAAAGTCAATTATTGT 948
Db 166295 AACTTCTCCTGTGGTGACAACTACAGTGTCTCTCAAGCAGTCTGTGAAAGTCAATTATTGT 166354

QY 949 TTCTGGAGCAACAGCACCAGAACTGTGTGAGTCACTTGTGACCAACTTGTGACCACTTGTGTGTC 1008
Db 166355 TTCTGGAGCAACAGCACCAGAACTGTGTGAGTCACTTGTGACCAACTTGTGACCACTTGTGTGTC 166414

QY 1009 AGTGGAGCAAAAGCTGGAGTGTGTGACACTTCTGTGGGCCAACTGTGTGACCACTTGTGTGTC 1068
Db 166415 AGTGGAGCAAAAGCTGGAGTGTGTGACACTTCTGTGGGCCAACTGTGTGACCACTTGTGTGTC 166474

QY 1069 AGGAACAACAGCTGGAACTGGTGTGCTTTCAGACTTCAAAACCCTTGTGACATCTGTGGC 1128
Db 166475 AGGAACAACAGCTGGAACTGGTGTGCTTTCAGACTTCAAAACCCTTGTGACATCTGTGGC 166534

QY 1129 AATACAGTGCACCGCTCTCACTGCAACCTGAAACCGAGTGTCTCTGGAACAGCAGT 1188
Db 166535 AATACAGTGCACCGCTCTCACTGCAACCTGAAACCGAGTGTCTCTGGAACAGCAGT 166594

QY 1189 AACACTGTCTTCCAGCAGTAACTTTTGGAGAACTTTCAGGTGCAGCTATTGTCTTCC 1248
Db 166595 AACACTGTCTTCCAGCAGTAACTTTTGGAGAACTTTCAGGTGCAGCTATTGTCTTCC 166654

QY 1249 ATCTGTGAAACCTGTGTGTCTTCTGCTGGGACCACATCTG -CAAGCCCTGTTATTGGGA 1307
Db 166655 ATCTGTGAAACCTGTGTGTCTTCTGCTGGGACCACATCTG -CAAGCCCTGTTATTGGGA 166713

QY 1308 CTCAGTTCAAATCAAACTTGGCCAGCGCGGCGCTGTCTTTCACACACAGCTGGGATTC 1367
Db 166714 CTCAGTTCAAATCAAACTTGGCCAGCGCGGCGCTGTCTTTCACACACAGCTGGGATTC 166773

QY 1368 CAACAGGCGCTTCAAGCAAGCACTATT 1395
Db 166774 CAACAGGCGCTTCAAGCAAGCACTATT 166801
```

```
RESULT 7
LOCUS 123466
DEFINITION Sequence 15 from patent US 5534410.
ACCESSION 123466
VERSION 123466.1 GI:1603336
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
```

```
Unclassified.
1 (bases 1 to 3603)
Tjian,R., Comai,L., Dynlacht,B.D., Hoey,T., Ruppert,S., Tanese,N.,
Wang,E. and Weinzierl,R.O.J.
TATA-binding protein associated factors drug screens
Patent: US 5534410-A 15 09-JUL-1996;
JOURNAL
FEATURES
source location/Qualifiers
BASE COUNT 969 a 961 c 830 g 843 t
ORIGIN
Query Match 13.2%; Score 338.2; DB 6; Length 3603;
Best Local Similarity 63.9%; Pred. No. 1.4e-77;
Matches 536; Conservative 0; Mismatches 288; Indels 15; Gaps 1;

QY 1586 TCTCTCACTTCAAGCATCTCCTACTCAGAAAAATAGATAAAGAGAAATGTAACATCATG 1645
Db 1377 TGTCTCGGCACACAGCTGCTGCACAGAAAAATAACTCAAGGAGCCTGGGGAGGTTTC 1436

QY 1646 CTTCCGAGATGAGGATGACATCAATGATGTGACTTCTATGGCAGGGGTCAACCTTAATGA 1705
Db 1437 GTTTCGGGAGCATGATGACATTAATGATGTGATCGATGCTGGAGTAAACTTGTCTAGA 1496

QY 1706 AGAAAAATGCCTGATCTTAGCAAAACTCTGAATTTGGTTGGCACACTTATTCAGTCAATG 1765
Db 1497 AGAAAGTGCAAGAAATATTAGCCAGAACTCTGAATTTGGTTGGCACACTTATTCAGTCAATG 1556

QY 1766 TAAAGATGAACCAATCTCTTTTATTGGAGCTCTCAAAAGAGAAATCTTAGACATTTGTTAA 1825
Db 1557 TAAAGATGAACCAATCTCTCTCCCAAGCGCTTTCAGAGAAAGAAATATTAGAAATAGGTAA 1616

QY 1826 AAAGCATGACATTTACAGAACTTAACTCTGATGCTGTGAACCTTGATCTCCCAAGCAACACA 1885
Db 1617 AAACATGTTATTAACGGAATTTACATCCAGATGTAGTAAGTTATCTATCATCATGCCGCA 1676

QY 1886 GGAACGACTAGGAGCGCTTCTAGAAAAACTGACTGCAATTTGCTCTGAGCATCGAATGACTAC 1945
Db 1677 ACAAGCTTACAGAAATCTTGTAGAGAAATATCAGAAACAGCTCAGCAGAGAAACTTTTC 1736

QY 1946 TTCAAGGCAAGTCAAAATTTACATCTTGTGTAGTATACAGGTCACAGGTCACAGCTCAAAATTTCT 2005
Db 1737 TTACAGGATGACACAGATATGAGCAGCGAGTGTAGCTCCGGCAGCTCAAGTTTTC 1796

QY 2006 TGAAGCTCGATCAATTTGGAGAGCAGAGAAAGGATTTTGAAGAAAGAGAAATCTTACT 2065
Db 1797 TGAACAGCTTGTCAAAATCGAAAAAGCAGAGGAGGATGAGCAGGCGGAGATCCCTGAT 1856

QY 2066 TAAGGAGCCAAAGAGTCTGTTCTTAATAAGAGATCCAGACAGCTGAGATTTAAAGCAGAA 2125
Db 1857 GAGGCGAGCAAAAGTCTCGGTCAAGCAAGAGAGATCCAGAACTTAAGGCTGAAACAGAA 1916

QY 2126 AGCCAAAGAGTTACAGCAATTTGGAACCTTGCACAGATACAGATAGACAGCTTAATCTCAC 2185
Db 1917 GGCNAAGGATGACGCAACAGGAACTTGGCAAAATGAGCAGCGGAGCCCAACCTCAC 1976

QY 2186 AGCTTCTTGCAGCTATTGGACCAAGAAAGAGAGACCACTAGA -----ATC 2230
Db 1977 AGCACTAGCAGCGATCGGGCCAGAAAAAGAGGAAAGTGGAGCTCTCGGGGCGCGGCTC 2036

QY 2231 TGGNAATTGAGGCTTAAAGACAACTTCTTCTGCTTGGGACATCCAGCCTGACAGCCAC 2290
Db 2037 AGGAGCAGAGGGGTCTGGGCGCGCGCTCAGTGTGTTCCAGGCACTCGGTGTCTGGAACCC 2096

QY 2291 CAAACAGTTGTCATCGTCAAGAAATCAGGAAATCTGCTCAGGACCTTGTATTTTGTAT 2350
Db 2097 CAGACATTTACGCGCAAAAGAAATCAGCGGGTCAACCTCAGGACCTCATATTTTGT 2156

QY 2351 GGAACAGGAACGGGAGATGAAGTATTCTCGAGCTCTATACCTGCGCCCTTCTGAAGTAC 2409
Db 2157 AGAAATGAACGTGAGACAAAGCCCATTTCACTGCTCTCTACAAAGCACTTCTCTTAAGTGAC 2215
```

[illegible]

Qy	2291	CAACACAGTTCATCGTCCAAAGAAATCACGAGAAATCTGCCTCAGGAGCTATGATATTTTGTAT	2350
Db	2097	CAGACAGTTCACGGGACAAAGAAATCACGGGGTCAACCTCAGGAGCCTCATATTTTGT	2156
Qy	2351	GGACACAGCGGAGATGAAGTATTCCTCGAGCTCTATACCTGCGCCCTCTCGAAGTGC	2409
Db	2157	AGAAATGAACGTGAGACAAAGCCATTCACCTGCTGCTCTACAAAGCATTCCTTAGTGAC	2215
RESULT	9		
LOCUS	HSU75308		
DEFINITION	Human TBP-associated factor (hTAFII130) mRNA, partial cds.		
ACCESSION	U75308		
VERSION	U75308.1	GI:1732072	
KEYWORDS	Homo sapiens.		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 4233)		
AUTHORS	Tanese, N., Saluja, D., Vassallo, M.F., Chen, J.L. and Admon, A.		
TITLE	Molecular cloning and analysis of two subunits of the human TFIID complex: hTAFII130 and hTAFII100		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 93 (24), 13611-13616 (1996)		
MEDLINE	97098442		
PUBMED	8942982		
REFERENCE	2 (bases 1 to 4233)		
AUTHORS	Tanese, N.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-OCT-1996) Microbiology, NYU Medical Center, 550 First Ave., New York, NY 10016, USA		
COMMENT	See Genbank Accession Number U75309 for hTAFII100; the 5' end is missing from the cDNA as this region is rich in GC residues; a few ambiguities remain in the cDNA sequence that may affect the reading frame near the 5' end; repeated efforts to obtain the full cDNA have been unsuccessful; numerous sequence runs have been performed to clarify the GC sequence ambiguities; the process is still ongoing.		
FEATURES	Location/Qualifiers		
source	1..4233		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/cell_line="HeLa"		
gene	1..4233		
	/gene="hTAFII130"		
CDS	<1..2844		
	/gene="hTAFII130"		
	/note="the C-terminal region is similar to Drosophila hTAFII10, Swiss-Prot Accession Number P47825; 130 kDa TAF subunit"		
	/codon_start=1		
	/product="TBP-associated factor"		
	/protein_id="AAC50901.1"		
	/db_xref="GI:1732073"		
	/translation="GRGLQQRGGREGSPAPPAAKLRPPPEGSAGACAPVPAATAAAGVGEPAAPAGKPNAGPAALARAAPGPGPGPGKPGKPAAGPAATLNGSAAALLNSHHAAPAGLVNGNPAALLPLPVPAPAAAPPPPPAPAPLARPCHGAPPPNALAVPPNPAQNGSAGAPAPAPAGGAGPGVCGPGGAAAAPAPKAESPVRVVAAPPAAGTAAASGAPASMSWIPTGMQGLSPAAVPPAPGPTGLPKGAAGAVTQSLSRTATTSGIRATLPTVLARLPQPNPTNQNFQPLPGVMVLVRSNGQLMLIPQQAQMAQAHAQRTSPAPVQISTVQAQPTIARQVPTTIIKOVSOAQTQVOPSAQTRSPGVQPOLVGGAAQTASLTATAVGTQPTVPGATTSSAATEWENYKKNFSLTLIKLASSKQSTETAAHYKELVQNLDDKIEADFTSLRYELNLSQPTLYLVFLKRLSKLQLTSDSAAFIQSQQPPPTTSQATLTALTAVLSSSVRTAGKTATVTSALQPVLSLTQPTGVGKGQPTPLVQQPKPALPQVQTVLTQTPVALRQPHRIMLTTPQQLIQTLQPLQPVVVKPALPGTKALSAVSAQAAAQKQLKPGGGSFRRDDINDVASMGVNLSEESARLATNSLVGLTTRSCDKDTFLLQAPLQRLLRIGKHHGTELHPDVSVSHATQORLQNLVEKISETAQOKNFYSYKDDRDYEQASDVRAQLKFFPQLDQIEQRDEQEREILMRAAKSRSGDEPQURLKQAKMOQQELMAGQRDANULTAAIGAPRRKKVDCPGPGSAGBGGSGVSPGSGVGTGTPQRTIRTRVYRDLRDLIFLCLENERETSHSLLYKAFLLK"		
BASE COUNT	1024 a 1276 c 1054 g 879 t		

BASE COUNT


```
QY 1766 TAAAGATGAACCAATTTCTTTTATTTGAGAGCTCTACAAAAGAGAAATCTTAGACATTTGGTAA 1825
Db 2595 TAAAGATGAACCAATTTCTCTCCCAAGCGCTTTGCAGAGAAGAAATATTAGAATAGGTAA 2654
QY 1826 AAAGCATGACATACAGAAATTAACCTGATCTGCTGTAAGTCTGCTCCCAAGCAACACA 1885
Db 2655 AAAACATGGTATACCGAATATACATCCAGATGATGTAAGTATGATATCATATGCCACGA 2714
QY 1886 GGAAGCATACGAGCGCTTCTAGAAAACTGACTGCAATTTGCTCAGCATCGAATGACTAC 1945
Db 2715 ACAAGGCTACAGAACTCTTAGAGAAATATACAGAAACAGCTCAGCAGAGAACTTTTC 2774
QY 1946 TTACAAGCAAGTCAAAATTTACATCTGTGTAGTATACCAAGTCAAGCTCAAAATTTCT 2005
Db 2775 TTACAAGGATGACCAAGATATGAGCAGCGAGTACGCTCCGGGCACAGCTCAAGTTTTT 2834
QY 2006 TGAAGAGCTGGATCAATTTGGAGNAGCAGAGAAAGGATTTGGAGAAAGAGAAATGTTACT 2065
Db 2835 TGAACAGCTTGATCAATCGAAAGCAGAGAGGATGAGCAGGAGGATCCTGAT 2894
QY 2066 TAAAGCAGCAAGAGTCTTCTAATAAAGAAAGATCCAGAACAGCTGAGATTTAAAGCAGAA 2125
Db 2895 GAGGCGAGCAAGTCTCGTCAAGACAGAGATCCAGAACAGTTAAGCTGNAACAGAA 2954
QY 2126 AGCCAAAGAGTTACAGCAATTTGGAATTTGCACAGATACAGCATAGACGCTAATCTCAC 2185
Db 2955 GGCNAAGGAGATGCAGCAACAGGAATGCGCAAAATGAGACAGCGGACGCCAACCTCAC 3014
QY 2186 AGCTCTTGAGCTATTTGGACCAAGGAGAGAGACCACTAGA-----ATC 2230
Db 3015 AGCAGTAGCAGCGATCGGGCCAGGAAAGAGAAAGTGGACTGTCTCGGGGCGCGGCTC 3074
QY 2231 TGGAAATGAGGGCTTAAAAGACAACCTCTTGTCTTGGGACATCCAGCTGACAGCCAC 2290
Db 3075 AGGAGCAGAGGGTGGCGCCCGCTCAGTGTCTCCAGCAGCTCGGGTGTCCGAACCCC 3134
QY 2291 CAACAGTTGCATCTGTCGAAGATACAGAGAAATCTGCCCTCAGGAGCTTGATATTTTGTAT 2350
Db 3135 CAGACAGTTTCACGCGACAAGAAATCACGGGGTCAACCTCAGGAGCTCATATTTTGTAT 3194
QY 2351 GGAAAGGAGGAGGAGATGAATTTCTCGAGCTCTATACCTGGCCCTCTGAAGTGA 2408
Db 3195 AGAAATGAACGTGAGACAAAGCAATTCAGTCTCTCAAGCAATTCCTTAAGTGA 3252

RESULT 11
AY038601
LOCUS
DEFINITION Mus musculus TATA-binding protein associated factor TAFII135 mRNA, partial cds.
ACCESSION AY038601
VERSION
KEYWORDS
SOURCE
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 2196)
AUTHORS Metsis M., Brunkhorst, A. and Neuman, T.
TITLE Cell Type Specific Expression of the TFIID Component TAFII135 in the Nervous System
JOURNAL Exp. Cell Res. (2001) In press
REFERENCE 2 (bases 1 to 2196)
AUTHORS Metsis M., Brunkhorst, A. and Neuman, T.
TITLE Direct Submissioin
JOURNAL Submitted (06-JUN-2001) Medical Chemistry and Biochemistry, Karolinska Institute, Retzius vag 1, Stockholm 1717, Sweden
FEATURES
source
1..2196
/organism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
/tissue_type="PO brain"
```

```
CDS
<1..1991
/codon_start=3
/product="TATA-binding protein associated factor TAFII135"
/protein_id="AAK94779.1"
/db_xref="GI:15375062"
/translation="ATLPTVLAPRLPQPKPNINQNPQLPPGMVLRSENGQLIMI
PQQAALQMAHAQAQAPQSTMAPRPATPTGAPVQISTVQAPGTPIIARQVPTTIIKQ
VSOAQTTVQPTTTLQRSQVQLVGGSAQSLGATATAVGTGTPORTVPQASTTST
AATETMENVKCKSLSTLILKASSGKOSTETAAVVKDLVNLGDKIEADFTSRLY
RELNSPQYLYPFLKRLPALRLQTPDSNAFIQQSQQQQPASQNTALTAVVLSSS
VORTAGKTAASVTSALQPPVLSLTQTVGKGQAPPTPLVITQQPKPGALIRPQVT
LTQTPWVALRQPHNRLMTLPOQIQNLQPVVVKVPLGTLKALSTVSAQAAQK
NKLKBPGGGSGFRDDDDINDVASMGVNLSEARILATNSLVLGTLTRCKDDTFLP
APLQSTLEIKKHGHTDCHPDVSVSHATQRLQNVVEKISETAQKNFSYKDDDR
YBQASDVRQLKFFEQDOIEXQRDEQEREILLMRAAKSRSEQPEQRLQKTKEM
QQQELAQNRQDRANLITALAIGFRKKRYDCTGPGSAGGSGFGAAVPGSGSGVTTPRQ
FTRQIRTRVNLRLDLIFCLENERETSHSLLLYKAFILVTG"
BASE COUNT 597 a 632 c 564 g 403 t
ORIGIN
Query Match 12.7%; Score 325.8; DB 10; Length 2196;
Best Local Similarity 63.1%; Pred. No. 2.6e-74;
Matches 526; Conservative 0; Mismatches 292; Indels 15; Gaps 1;
QY 1584 ATTCTGTCACTTCAAGCATCTCTACTCAGAAAAATAGATAAAGAGAAATGTAAACATCA 1643
Db 1146 ACTGTATCTGCACAACGAGCTGCTGCCAGAAAAACAACTCAAGGAGCGGGGGGAGGC 1205
QY 1644 TGTCTCCGAGATGAGGATGACATCAATGATGTGACTTCTATGGCAGGGGTCAACCTTAAT 1703
Db 1206 TCTTCCGGGATGATGACATCAATGATGTTGCGCTCAATGCTGGAGTGAACCTGTCT 1265
QY 1704 GAAGAAATAGCCCTGCATCTTAGCAACAACTCTGAATTTGTTGCGACACTCATTCAGTCA 1763
Db 1266 GAAGAAAGTGCAGAGATATTGCCACAACTCTGAGTTGTGGGAGAGTTGACCCGATCC 1325
QY 1764 TGTAAAGATGAACCAATTTCTTTTATTGGAGCTCTACAAAAGAGAAATCTTAGACATTTGT 1823
Db 1326 TGTAAAGACGACACCTTCTCTCTCCCGCCCTTTTCAGAGGAGACACTGGAGATAGGT 1385
QY 1824 AAAAGCATGACATTTACAGAACTTAATCTGTGCTGTGNACTTGATCTCCCAAGCAACA 1883
Db 1386 AAAAGCATGCGCATACGAGATGCTCCATCCAGATGTAGTACGTTACGTGTCTCATGCCACA 1445
QY 1884 CAGGAAGCATACGAGCGCTTCTAGAAAACTGACTTGCATTTCTCAGCATCAATGACT 1943
Db 1446 CAGCAAGGCTGCAGAAATGTTGTAGAAAAATAATCTGAAACGGCTCAGCAGAAAACTTC 1505
QY 1944 ACTTAAAGCAAGTGAATAATACATCTCTGTGTAGTATACCAAGTCAAGCTCAAAATTT 2003
Db 1506 TCTTATAAGGACGACGACAGATACGAGCAAGCGAGTGTCTCCGGCTCAGCTCAAAATTC 1565
QY 2004 CTTGAAAAGCTGGATCAATTTGGAGAGCAGAGAAAGGATTTGGAAGAAAGAGAAATGTTA 2063
Db 1566 TTTGAAACAATGGATCAGATTTGAAAAGCAGAGAGAAAGATGAGCAGGAGGAGAGATCTTA 1625
QY 2064 CTTAAGCGACCAAGAGTCTGTTCTTAATAAGAGATCCAGAACAGCTGAGATTAAGACGAG 2123
Db 1626 ATGCGGCGCAGGAGTCTCGATCCACAGAAAGAACCCAGAGCAATTAAGTTTGAAGCAG 1685
QY 2124 AAAGCCAAAGAGTTACAGCAATTTGGAATTTGACAGATACAGCATAGAGACGCTAATCTC 2183
Db 1686 AAAACAAAAGAGATGACGAGCAGGAGCTGGCACAGATGCGACACAGATGCCAACCTC 1745
QY 2184 ACAGCTCTTTGAGCTATTGGACCAAGAGAGAGACCA-----CTAGAA 2228
Db 1746 ACTGCACTGGCAGCATCGGGCCCGAGGAAAGAGAGAAAGTGGACTGCACTGGGCCAGGA 1805
QY 2229 TCTGGAATTTGAGGGCTTAAAGACAACTTCTTCTGCTTCTGGGACATCCAGCTCGACGCC 2288
Db 1806 TCAGGAGCAGAGGCTCAGCGCCGGCGGCGAGCAGTCCAGCGGGGTCCGCGTGGGACCC 1865
QY 2289 ACCAAACAGTTGATCGTCCAAAGAAATCACAGAAATCTGCTTCAGGAGCTTGATATTTTGT 2348
```

Db 1866 CCCAGACAGTTACACGCGCAAGATCACCGGTCACCTCAGGACCTCATATTTGT 1925
QY 2349 ATGGAACAGGAGGAGAGTATTCGAGCTCTATACCTGGCCCTTCT 2401
Db 1926 TTGAAGACGAGGCTGAGACAGCAATTCCTGTTGCTCTACAAAGCATTCCT 1978

RESULT 12
AP001197/c
LOCUS
DEFINITION
Homo sapiens chromosome 18 clone RP11-775B10 map 18q11.2, WORKING
DRAFT SEQUENCE, 30 unordered pieces.
ACCESSION
AP001197
VERSION
AP001197.2 GI:81117643
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens DNA, clone:RP11-775B10.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 17422)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 174,222 genomic DNA of 18q11.2
Published Only in DataBase (2000)
2 (bases 1 to 17422)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (18-FEB-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 30, 2000 this sequence version replaced gi:7023969.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-775B10
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 151996 bases at least Q40
Consensus quality: 162600 bases at least Q30
Consensus quality: 167916 bases at least Q20
Insert size: 171322; sum-of-contigs
Quality coverage: 4.35x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 30 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1 16728 contig of 16728 bp in length
16829 31270 contig of 14442 bp in length
31371 45555 contig of 14185 bp in length
45656 59177 contig of 13522 bp in length
59278 70338 contig of 11061 bp in length
70439 80210 contig of 9772 bp in length
80311 89495 contig of 9185 bp in length
89596 96401 contig of 6806 bp in length
102566 109526 contig of 6861 bp in length
109627 115496 contig of 5870 bp in length
115597 121502 contig of 5906 bp in length

121603 126153 contig of 4551 bp in length
126254 131007 contig of 4754 bp in length
131108 135231 contig of 4124 bp in length
135332 138844 contig of 3513 bp in length
138945 142268 contig of 3324 bp in length
142369 146076 contig of 3708 bp in length
146177 149651 contig of 3475 bp in length
149752 152991 contig of 3240 bp in length
153092 157112 contig of 4021 bp in length
157213 159865 contig of 2753 bp in length
160066 161994 contig of 1439 bp in length
162095 163533 contig of 1929 bp in length
163634 165427 contig of 1794 bp in length
165528 167111 contig of 1584 bp in length
167212 169213 contig of 2002 bp in length
169314 170970 contig of 1657 bp in length
171071 173121 contig of 2051 bp in length
173222 174222 contig of 1001 bp in length

Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently consists of 30 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 16728: contig of 16728 bp in length
* 16729 16828: gap of 100 bp
* 16829 31270: contig of 14442 bp in length
* 31271 31370: gap of 100 bp
* 31371 45555: contig of 14185 bp in length
* 45556 59177: gap of 100 bp
* 59178 59277: contig of 13522 bp in length
* 59278 70338: contig of 11061 bp in length
* 70339 70438: gap of 100 bp
* 70439 80210: contig of 9772 bp in length
* 80211 80310: gap of 100 bp
* 80311 89495: contig of 9185 bp in length
* 89496 89595: gap of 100 bp
* 89596 96401: contig of 6806 bp in length
* 96402 96501: gap of 100 bp
* 96502 102565: contig of 6064 bp in length
* 102566 102665: gap of 100 bp
* 102666 109526: contig of 6861 bp in length
* 109527 109626: gap of 100 bp
* 109627 115496: contig of 5870 bp in length
* 115497 115596: gap of 100 bp
* 115597 121502: contig of 5906 bp in length
* 121503 121602: gap of 100 bp
* 121603 126153: contig of 4551 bp in length
* 126154 126253: gap of 100 bp
* 126254 131007: contig of 4754 bp in length
* 131008 131107: gap of 100 bp
* 131108 135231: contig of 4124 bp in length
* 135232 135331: gap of 100 bp
* 135332 138844: contig of 3513 bp in length
* 138845 138944: gap of 100 bp
* 138945 142268: contig of 3324 bp in length
* 142269 142368: gap of 100 bp
* 142369 146076: contig of 3708 bp in length
* 146077 146176: gap of 100 bp
* 146177 149651: contig of 3475 bp in length
* 149652 149751: gap of 100 bp
* 149752 152991: contig of 3240 bp in length
* 152992 153091: gap of 100 bp
* 153092 157112: contig of 4021 bp in length
* 157113 157212: gap of 100 bp
* 157213 159865: contig of 2753 bp in length
* 159866 160065: gap of 100 bp
* 160066 161994: contig of 1439 bp in length
* 161995 162094: gap of 100 bp

COMMENT

* 162095 163533: contig of 1439 bp in length
 * 163534 163633: gap of 100 bp
 * 163634 163647: contig of 1794 bp in length
 * 163648 165527: gap of 100 bp
 * 165528 167111: contig of 1584 bp in length
 * 167112 167211: gap of 100 bp
 * 167212 169213: contig of 2002 bp in length
 * 169214 169313: gap of 100 bp
 * 169314 170970: contig of 1657 bp in length
 * 170971 171070: gap of 100 bp
 * 171071 173121: contig of 2051 bp in length
 * 173122 173221: gap of 100 bp
 * 173222 174222: contig of 1001 bp in length.

FEATURES

source

Location/Qualifiers
 1. .174222
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18q11.2"
 /clone="RP11-775B10"

misc_feature

1. .16728
 /note="assembly_fragment"

misc_feature

16829. .31270
 /note="assembly_fragment"

misc_feature

31371. .45555
 /note="assembly_fragment"

misc_feature

45656. .59177
 /note="assembly_fragment"

misc_feature

59278. .70338
 /note="assembly_fragment"

misc_feature

70439. .80210
 /note="assembly_fragment"

misc_feature

80311. .89495
 /note="assembly_fragment"

misc_feature

89596. .96401
 /note="assembly_fragment"

misc_feature

96502. .102565
 /note="assembly_fragment"

misc_feature

102866. .109526
 /note="assembly_fragment"

misc_feature

109627. .115496
 /note="assembly_fragment"

misc_feature

115597. .121502
 /note="assembly_fragment"

misc_feature

121603. .126153
 /note="assembly_fragment"

misc_feature

126254. .131007
 /note="assembly_fragment"

misc_feature

131108. .135231
 /note="assembly_fragment"

misc_feature

135332. .138844
 /note="assembly_fragment"

misc_feature

138945. .142268
 /note="assembly_fragment"

misc_feature

142369. .146076
 /note="assembly_fragment"

misc_feature

146177. .149651
 /note="assembly_fragment"

misc_feature

149752. .152991
 /note="assembly_fragment"

misc_feature

153092. .157112
 /note="assembly_fragment"

misc_feature

157213. .159965
 /note="assembly_fragment"

Query Match

Best Local Similarity 11.5%; Score 294.6; DB 2; Length 174222;

Matches 319; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

Qy 2236 TTGAGGCTTAAAGACACCTTCTTGGCTTGGGACATCCAGCTGACAGCCACCAAC 2295

|| |||||

Db 174026 TTGAGGCTTAAAGACACCTTCTTGGCTTGGGACATCCAGCTGACAGCCACCAAC 173967

|| |||||

QY 2296

AGTTGTCATCGTCCGAAGATCAGGAAATCGCTCAGGGACTTCATATTTTGTATGAAC 2355

Db 173966

AGTTGTCATCGTCCGAAGATCAGGAAATCGCTCAGGGACTTCATATTTTGTATGAAC 173907

QY 2356

AGGAACGGGAGATGAAGTATTCCTCGAGCTCTATACCTGGCCCTCTCTGAAGTACCCTCC 2415

Db 173906

AGGAACGGGAGATGAAGTATTCCTCGAGCTCTATACCTGGCCCTCTCTGAAGTACCCTCC 173847

QY 2416

ACTCTTCCATCCACATCCTTGCTATTCTGCGCAAGAGACACAAAGCATTTCTGCACCT 2475

Db 173846

ACTCTTCCATCCAGATCCTTGCTATTCTGCGCAAGAGACACAAAGCATTTCTGCACCT 173787

QY 2476

GTCTGAAATTTCAATTTCTGGAATAA-CACCAACATGAAGAGCATTTAGGATT 2534

Db 173786

GTCTGAAATTTCAATTTCTGGAATAA-TACCAACATGAAGAGCATTTAGGATT 173727

QY 2535

AG-AACCTTTATTAACCTTTACCTAT 2558

Db 173726

AGAAACTTTATTAACTCTTACCTAT 173702

RESULT 13

AP001096

LOCUS

DEFINITION

AP001096

ACCESSION

AP001096.4 GI:9188475

VERSION

HTG: HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Jul 14, 2000 this sequence version replaced gi:8117851.

Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: <http://hgp.gsc.riken.go.jp/>Contact: hattori@gsc.riken.go.jp

----- Project Information

Center project name: HumDraft18

Center clone name: RP11-813H9

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 172065 bases at least Q40

Consensus quality: 173442 bases at least Q30

Consensus quality: 174017 bases at least Q20

Insert size: 174453; sum-of-contigs

Quality coverage: 10.99x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of

12 contigs. The true order of the pieces is not known and their

order in this sequence record is arbitrary. Gaps between the

contigs are represented as runs N, but the exact sizes of the gaps

are unknown. This record will be updated with the finished sequence

as soon as it is available and the accession number will be

preserved

1 65994 contig of 65994 bp in length
66095 80769 contig of 14675 bp in length
80870 95866 contig of 14717 bp in length
95867 108410 contig of 12724 bp in length
108511 122077 contig of 13567 bp in length
122178 135900 contig of 13723 bp in length
136001 149231 contig of 13231 bp in length
149332 158870 contig of 9539 bp in length
158971 165511 contig of 6541 bp in length
165612 171812 contig of 6201 bp in length
171913 173894 contig of 1982 bp in length
173995 175553 contig of 1559 bp in length.
* NOTE: This is a 'working draft' sequence. It currently consists of 12 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
* 1 65994: contig of 65994 bp in length
* 65995 66094: gap of 100 bp
* 66095 80769: contig of 14675 bp in length
* 80770 80869: gap of 100 bp
* 80870 95866: contig of 14717 bp in length
* 95867 95866: gap of 100 bp
* 95867 108410: contig of 12724 bp in length
* 108411 108510: gap of 100 bp
* 108511 122077: contig of 13567 bp in length
* 122078 122177: gap of 100 bp
* 122178 135900: contig of 13723 bp in length
* 135901 136000: gap of 100 bp
* 136001 149231: contig of 13231 bp in length
* 149232 149331: gap of 100 bp
* 149332 158870: contig of 9539 bp in length
* 158871 158970: gap of 100 bp
* 158971 165511: contig of 6541 bp in length
* 165512 165611: gap of 100 bp
* 165612 171812: contig of 6201 bp in length
* 171813 171912: gap of 100 bp
* 171913 173894: contig of 1982 bp in length
* 173895 173994: gap of 100 bp
* 173995 175553: contig of 1559 bp in length.
FEATURES
source
1. .175553
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q11.2"
/clone="RP11-813H9"
misc_feature
1. .65994
/note="assembly_fragment"
66095. .80769
/note="assembly_fragment"
80870. .95866
/note="assembly_fragment"
95867. .108410
/note="assembly_fragment clone_end:T7 vector_side:left"
108511. .122077
/note="assembly_fragment"
122178. .135900
/note="assembly_fragment"
136001. .149231
/note="assembly_fragment"
149332. .158870
/note="assembly_fragment"
158971. .165511
/note="assembly_fragment clone_end:SP6 vector_side:right"
165612. .171812
/note="assembly_fragment"
171913. .173894
/note="assembly_fragment"
173995. .175553

BASE COUNT 50929 a 36376 c 36886 g 50262 t 1100 others
ORIGIN
Query Match 11.5%; Score 294.6; DB 2; Length 175553;
Best Local Similarity 98.2%; Pred. No. 3.2e-66;
Matches 319; Conservative 0; Mismatches 4; Indels 2; Gaps 2;
QY 2236 TTGAGGCTTAAAGACAACTTCTTGGTCTTGGGACATCCAGCTGACAGCCACCAAC 2295
DB 104707 TTTAGGCTTAAAGACAACTTCTTGGTCTTGGGACATCCAGCTGACAGCCACCAAC 104766
QY 2296 AGTTGCATCGTCAAGATCAGGAGATCTCGCTCAGGACTTGTATTTGTATGGAC 2355
DB 104767 AGTTGCATCGTCAAGATCAGGAGATCTCGCTCAGGACTTGTATTTGTATGGAC 104826
QY 2356 AGGAACGGGAGATGAAGTATTTCTCGAGCTCTATACCTGGCCCTTCTGAAGTGACCACTCC 2415
DB 104827 AGGAACGGGAGATGAAGTATTTCTCGAGCTCTATACCTGGCCCTTCTGAAGTGACCACTCC 104886
QY 2416 ACTCTTCATCCACATCCTTGTCTATTTACTGCCAAGAACACAAAGCATTTGTGCACT 2475
DB 104887 ACTCTTCATCCACATCCTTGTCTATTTACTGCCAAGAACACAAAGCATTTGTGCACT 104946
QY 2476 GTCTGGAATTTCAATTTCTGGAATTAACACCAACATGAAGAGCATTTGTACGATT 2534
DB 104947 GTCTGGAATTTCAATTTCTGGAATTAACACCAACATGAAGAGCATTTGTACGATT 105006
QY 2535 AG-AACTTTATTAACTCTTACCTAT 2558
DB 105007 AGAATTTTATTAACTCTTACCTAT 105031
RESULT 14
AC007996
LOCUS 182884 bp DNA linear PRI 22-APR-2002
DEFINITION Homo sapiens chromosome 18, clone RP11-9E17, complete sequence.
ACCESSION AC007996
VERSION AC007996.11 GI:20128705
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 182884)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatsis,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwen,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (08-JUL-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 182884)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,

Farro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meidrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 182884)

Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Farro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hago, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meidrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (22-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 10, 2002 this sequence version replaced gi:16041571.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: L954
Center clone name: 9_E17

FEATURES

source
1..182884
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-9E17"
/clone_lib="RP11-9E17 Human Male BAC"
complement(711..1245)
/rpt_family="L2"
complement(2401..2496)
/rpt_family="MIR"
2522..2595
/rpt_family="L2"

repeat_region
complement(711..1245)
/rpt_family="L2"
complement(2401..2496)
/rpt_family="MIR"
2522..2595
/rpt_family="L2"

repeat_region
complement(4047..4535)
/rpt_family="LIME3A"
complement(4697..4951)
/rpt_family="AluSx"
complement(6008..6119)
/rpt_family="L2"
6391..6431
/rpt_family="MIR"
complement(6432..6737)
/rpt_family="AluSx"
6738..6916
/rpt_family="MIR"
complement(7019..7142)
/rpt_family="LIME10"
complement(7157..7262)
/rpt_family="L2"
9285..9308
/rpt_family="AT_rich"
9584..9620
/rpt_family="AT_rich"
9772..9897
/rpt_family="Charlie8"
9979..10250
/rpt_family="AluJo"
10254..10542
/rpt_family="AluSg"
10544..10568
/rpt_family="(TG)n"
complement(10758..11031)
/rpt_family="AluJo"
complement(11044..11392)
/rpt_family="AluSx"
12404..12555
/rpt_family="MIR"
complement(12537..12575)
/rpt_family="MIR"
complement(12608..12877)
/rpt_family="Tigger3(Golem)"
complement(12962..13263)
/rpt_family="AluSx"
complement(14519..14584)
/rpt_family="LIME4A"
complement(15397..15834)
/rpt_family="MLT1C"
complement(17303..17387)
/rpt_family="L3"
complement(17510..17606)
/rpt_family="MIR"
17613..17795
/rpt_family="MLT1J2"
complement(17796..18091)
/rpt_family="AluJb"
18092..18335
/rpt_family="MLT1J2"
complement(18345..18495)
/rpt_family="MIR"
complement(18522..18778)
/rpt_family="L3"
18861..18938
/rpt_family="MER5A"
complement(18939..19110)
/rpt_family="AluSg/x"
19112..19136
/rpt_family="(TG)n"
19133..19148
/note="Single clone coverage"
complement(19137..19182)
/rpt_family="LTR66"
19183..19283
/rpt_family="(TA)n"
complement(19284..19417)


```
* 26141 26240: gap of 100 bp
* 26241 36929: contig of 10689 bp in length
* 36930 37029: gap of 100 bp
* 37030 48911: contig of 11882 bp in length
* 48912 49011: gap of 100 bp
* 49012 57744: contig of 8733 bp in length
* 57745 57844: gap of 100 bp
* 57845 68839: contig of 10995 bp in length
* 68840 68939: gap of 100 bp
* 68940 77569: contig of 8629 bp in length
* 77569 77668: gap of 100 bp
* 77669 87757: contig of 9989 bp in length
* 87758 87757: gap of 100 bp
* 87758 93985: contig of 6228 bp in length
* 93986 94085: gap of 100 bp
* 94086 100964: contig of 6879 bp in length
* 100965 101064: gap of 100 bp
* 101065 107297: contig of 6233 bp in length
* 107298 107397: gap of 100 bp
* 107398 113948: contig of 6551 bp in length
* 113949 114048: gap of 100 bp
* 114049 120683: contig of 6635 bp in length
* 120684 120783: gap of 100 bp
* 120784 126925: contig of 6142 bp in length
* 126926 127025: gap of 100 bp
* 127026 130951: contig of 3926 bp in length
* 130952 131051: gap of 100 bp
* 131052 134620: contig of 3569 bp in length
* 134621 134720: gap of 100 bp
* 134721 138171: contig of 3451 bp in length
* 138172 138271: gap of 100 bp
* 138272 143486: contig of 5215 bp in length
* 143487 143586: gap of 100 bp
* 143587 148147: contig of 4561 bp in length
* 148148 148247: gap of 100 bp
* 148248 152197: contig of 3950 bp in length
* 152198 152297: gap of 100 bp
* 152298 155994: contig of 3697 bp in length
* 155995 156094: gap of 100 bp
* 156095 159446: contig of 3352 bp in length
* 159447 159546: gap of 100 bp
* 159547 162952: contig of 3406 bp in length
* 162953 163052: gap of 100 bp
* 163053 165356: contig of 2504 bp in length
* 165357 165656: gap of 100 bp
* 165657 168139: contig of 2483 bp in length
* 168140 168239: gap of 100 bp
* 168240 170861: contig of 2622 bp in length
* 170862 170961: gap of 100 bp
* 170962 173710: contig of 2749 bp in length
* 173711 173810: gap of 100 bp
* 173811 175823: contig of 2013 bp in length
* 175824 175923: gap of 100 bp
* 175924 177259: contig of 1336 bp in length
* 177260 177359: gap of 100 bp
* 177360 178684: contig of 1325 bp in length
* 178685 178784: gap of 100 bp
* 178785 180676: contig of 1892 bp in length
* 180677 180776: gap of 100 bp
* 180777 182266: contig of 1490 bp in length
* 182267 182366: gap of 100 bp
* 182367 184572: contig of 2206 bp in length
* 184573 184672: gap of 100 bp
* 184673 186271: contig of 1599 bp in length
* 186272 186371: gap of 100 bp
* 186372 188088: contig of 1717 bp in length
* 188089 188188: gap of 100 bp
* 188189 189313: contig of 1125 bp in length
* 189314 189413: gap of 100 bp
* 189414 190585: contig of 1172 bp in length
* 190586 190685: gap of 100 bp
* 190686 190782: contig of 97 bp in length.
```

FEATURES

Location/Qualifiers

source

```
1. .190782
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q11.2"
/clone="RP11-802C10"
1. .14910
/note="assembly_fragment"
15011. .26140
/note="assembly_fragment"
26241. .36929
/note="assembly_fragment"
37030. .48911
/note="assembly_fragment"
49012. .57744
/note="assembly_fragment"
57845. .68839
/note="assembly_fragment"
68940. .77568
/note="assembly_fragment"
77669. .87657
/note="assembly_fragment"
87758. .93985
/note="assembly_fragment clone_end:T7 vector_side:left"
94086. .100964
/note="assembly_fragment"
101065. .107297
/note="assembly_fragment"
107398. .113948

Query Match 11.5%; Score 294.6; DB 2; Length 190782;
Best Local Similarity 98.2%; Pred. No. 3.1e-66;
Matches 319; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 2236 TTGAGGGCTTAAAGACAACTTCTTGGCTCTGGGACATCCAGCCTGACAGCCACCAAC 2295
|| |||||
Db 153393 TTAGGGCTTAAAGACAACTTCTTGGCTCTGGGACATCCAGCCTGACAGCCACCAAC 153334

QY 2296 AGTTGCATCGTCCAGAAATCAGGAAATCTGCTCAGGACTTGATATTTGTATGGAAC 2355
|| |||||
Db 153333 AGTTGCATCGTCCAGAAATCAGGAAATCTGCTCAGGACTTGATATTTGTATGGAAC 153274

QY 2356 AGAACGGGAGATGAAGTATTCGAGCTCTATACCTGGCCCTTCTGAAAGTACCCTCC 2415
|| |||||
Db 153273 AGAACGGGAGATGAAGTATTCGAGCTCTATACCTGGCCCTTCTGAAAGTACCCTCC 153214

QY 2416 ACTCTCCATCCACATCCTTGCTATTACTGCAAAAGACACAAAGCATTTGTCACCT 2475
|| |||||
Db 153213 ACTCTCCATCCAGATCCTTGCTATTACTGCAAAAGACACAAAGCATTTGTCACCT 153154

QY 2476 GTCCTGAAATTTCAATTTCTGGAATAA-CACCAACATGAAAGAGCATTTGTTAGGATT 2534
|| |||||
Db 153153 GTCCTGAAATTTCAATTTCTGGAATAAATACCAACATGAAAGAGCATTTGTTAGGATT 153094

QY 2535 AG-ACTTTTAACTCTTACCTAT 2558
|| |||||
Db 153093 AGAACTTTTAACTCTTACCTAT 153069
```

Search completed: February 17, 2003, 02:40:41
Job time : 9051 secs

GenCore version 5.1.3
Copyright (c) 1993-2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2003, 22:29:40 ; Search time 6475 Seconds
(without alignments)
11497.294 Million cell updates/sec

Title: US-09-763-909-1

Perfect score: 2558

Sequence: 1'gggacctgtgacaaagt.....ctttattaaactttaacctat 2558

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_mu:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_nam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2534	99.1	2556	6	I81227
2	2534	99.1	2556	9	HSTAFII
3	1145.2	44.8	182805	9	AC017007 Homo sapi
4	580	22.7	119081	2	AC121320 Homo sapi
5	580	22.7	169240	9	AC022089 Homo sapi
6	580	22.7	190782	2	AP002752 Homo sapi
7	338.2	13.2	3603	6	I23466
8	338.2	13.2	3603	6	I45759
9	338.2	13.2	4233	9	HSU75308
10	337.2	13.2	3252	9	HSTAFII13
11	325.8	12.7	2196	10	AY038601
12	294.6	11.5	174222	2	AP001197
13	294.6	11.5	175553	2	AP001096
14	294.6	11.5	182884	9	AC007996 Homo sapi
15	294.6	11.5	190782	2	AP002752
16	148.6	5.8	333300	2	AC125091
17	148	5.8	16903	2	AC016839
18	148	5.8	61882	2	AC118057
19	147	5.7	277892	2	AC125105
20	146.6	5.7	205466	2	AC127767
21	143.8	5.6	3535	3	S63550
22	142.2	5.6	3745	3	AY069807
23	142.2	5.6	4615	3	DROTAFL10X
24	142.2	5.6	4615	6	I23459
25	142.2	5.6	4615	6	I45752
26	137.6	5.4	205466	2	AC127767
27	124.4	4.9	62974	2	AC118056
28	121.8	4.8	277892	2	AC125105
29	112.4	4.4	158509	2	AC114177
30	109.8	4.3	62974	2	AC118056
31	99	3.9	158509	2	AC114177
32	98.4	3.8	191918	2	AC102692
33	94.6	3.7	121982	9	HS1107C24
34	92	3.6	333300	2	AC125091
35	91.2	3.6	64035	2	AC119227
36	89.6	2.5	194426	2	AL663067
37	71	2.8	112277	2	AC113339
38	62.4	2.4	2147	9	AK097744
39	58.6	2.3	59865	2	AC014422
40	58.6	2.3	168479	3	AC093454
41	58.6	2.3	176056	3	AC010066
42	58.6	2.3	283821	3	AE003528
43	57.6	2.3	7218	6	I66494
44	56.8	2.2	126712	9	AL137077
45	56.4	2.2	201470	2	AC113649

ALIGNMENTS

RESULT 1
I81227
LOCUS I81227 2556 bp DNA linear PAT 10-JUN-1998
DEFINITION Sequence 1 from patent US 5710025.
ACCESSION I81227
VERSION I81227.1 GI:3209517
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2556)
AUTHORS Dikstein,R. and Tjian,R.
TITLE Cell-type specific transcription factor
JOURNAL Patent: US 5710025-A 1 20-JAN-1998;
FEATURES Location/Qualifiers

This Page Blank (uspto)